SIAM J. NUMER. ANAL.
Vol. 60, No. 5, pp. 2448–2481

ON ENERGY LAWS AND STABILITY OF RUNGE–KUTTA METHODS FOR LINEAR SEMINEGATIVE PROBLEMS*

ZHENG SUN†, YUANZHE WEI‡, AND KAILIANG WU§

Abstract. This paper presents a systematic theoretical framework to derive the energy identities of *general implicit and explicit* Runge–Kutta (RK) methods for linear seminegative systems. It generalizes the stability analysis of only *explicit* RK methods in [Z. Sun and C.-W. Shu, *SIAM J. Numer. Anal.*, 57 (2019), pp. 1158–1182]. The established energy identities provide a precise characterization on whether and how the energy dissipates in the RK discretization, thereby leading to weak and strong stability criteria of RK methods. Furthermore, we discover a unified energy identity for all the diagonal Padé approximations, based on an analytical Cholesky type decomposition of a class of symmetric matrices. The structure of the matrices is very complicated, rendering the discovery of the unified energy identity and the proof of the decomposition highly challenging. Our proofs involve the construction of technical combinatorial identities and novel techniques from the theory of hypergeometric series. Our framework is motivated by a discrete analogue of integration by parts technique and a series expansion of the continuous energy law. In some special cases, our analyses establish a close connection between the continuous and discrete energy laws, enhancing our understanding of their intrinsic mechanisms. Several specific examples of implicit methods are given to illustrate the discrete energy laws. A few numerical examples further confirm the theoretical properties.

Key words. Runge–Kutta methods, energy laws, L^2-stability, Padé approximations, energy method

MSC codes. 65M12, 65L06, 65L20, 15A23

DOI. 10.1137/22M1472218

1. Introduction.

This paper is concerned with the autonomous linear seminegative differential systems in a general form:

$$\frac{\mathrm{d}}{\mathrm{d}t}u = Lu, \quad u = u(t) \in L^2([0,T];V), \tag{1.1}$$

where V is a finite or infinite dimensional real Hilbert space equipped with the inner product $\langle\cdot,\cdot\rangle$ and the induced norm $\|\cdot\|$, and L is a bounded linear seminegative operator satisfying $\langle Lv, v\rangle \le 0$ for all $v \in V$. (The operator L is not necessarily normal, namely, it may not commute with its adjoint.) A typical example of (1.1) is the linear seminegative ordinary differential equations (ODEs) with $V = \mathbb{R}^{N_d}$, $\langle\cdot,\cdot\rangle$ being the standard l^2 inner product, and the operator L being a seminegative $N_d \times N_d$ real constant matrix. Such ODEs may also arise from suitable semidiscrete schemes

*Received by the editors January 18, 2022; accepted for publication (in revised form) June 16, 2022; published electronically September 13, 2022.
https://doi.org/10.1137/22M1472218

Funding: The work of the first author was partially supported by the NSF grant DMS-2208391. The work of the third author was partially supported by the National Natural Science Foundation of China grant 12171227.

†Department of Mathematics, The University of Alabama, Tuscaloosa, AL 35487 USA (zsun30@ua.edu).

‡Department of Mathematics, Southern University of Science and Technology, Shenzhen, Guangdong 518055, People's Republic of China (weiyz2019@mail.sustech.edu.cn).

§Corresponding author. Department of Mathematics & SUSTech International Center for Mathematics, Southern University of Science and Technology, and National Center for Applied Mathematics Shenzhen (NCAMS), Shenzhen, Guangdong 518055, People's Republic of China (wukl@sustech.edu.cn).

for some linear partial differential equations (PDEs), such as linear hyperbolic or convection-diffusion equations, etc. The seminegative operator L induces a semi-inner-product $[\cdot,\cdot]$ on V defined by

$$[w,v] := -\langle Lw, v\rangle - \langle w, Lv\rangle . \tag{1.2}$$

The corresponding seminorm is denoted as $[\![v]\!] := \sqrt{[v,v]}$. Then it can be seen that the system (1.1) admits the following energy dissipation law:

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u\|^2 = \left\langle \frac{\mathrm{d}}{\mathrm{d}t}u, u\right\rangle + \left\langle u, \frac{\mathrm{d}}{\mathrm{d}t}u\right\rangle = \langle Lu, u\rangle + \langle u, Lu\rangle = -[\![u]\!]^2 \le 0. \tag{1.3}$$

Furthermore, if we integrate (1.3) in time from t^n to $t^{n+1} := t^n + \tau$ with $\tau > 0$, then it yields

$$\left\|u(t^{n+1})\right\|^2 - \|u(t^n)\|^2 = -\int_0^\tau [\![u(t^n+\widehat{\tau})]\!]^2 \,\mathrm{d}\widehat{\tau} \le 0. \tag{1.4}$$

The Runge–Kutta (RK) methods are widely used in temporal discretization for the approximate solutions of ODEs and time-dependent PDEs. In this paper, we discretize system (1.1) with RK methods, and we wish to establish a systematic framework to study how the energy law (1.4) is approximated in generic RK discretization. The discrete energy laws are important and helpful for further understanding the stability of RK methods, which is a classical topic in numerical analysis. Over the past decades, rich mathematical theories on the stability of RK methods have been developed both in the ODE settings (see [39, Chapter IV], [5, Chapter 3], and references therein) and in the context of numerical PDEs (see [3, 44, 8, 38, 43, 42, 9] and references therein).

One classical way to analyze the stability of RK methods is through the eigenvalue analysis, which typically focuses on the scalar ODE $\frac{\mathrm{d}}{\mathrm{d}t}u = \lambda u$ with a complex constant λ. Specifically, an RK method applied to this scalar ODE reduces to the iteration $u^{n+1} = \mathcal{R}(\tau\lambda)u^n$ and the stability criterion is then imposed as $|\mathcal{R}(\tau\lambda)| \le 1$. In the special case that the stability region (where $|\mathcal{R}(\tau\lambda)| \le 1$ holds) contains the left complex plane, the methods are called A-stable [7]. It is noted that for an A-stable RK method, the unconditional stability for the scalar equation implies the L^2-stability for the linear seminegative system (1.1) in the sense that $\left\|u^{n+1}\right\| \le \|u^n\|$. A proof of this implication was given in [39, Chapter IV.11, pp. 179–180] based on a lemma by von Neumann [25]; see also [14]. However, for the RK methods that are not A-stable, special attention should be paid when extending the analysis from the scalar equation to the ODE system (1.1). If the operator L in (1.1) is normal, namely, it commutes with its adjoint, then the system (1.1) can be unitarily diagonalized into decoupled scalar equations. In this case, the eigenvalue analysis will provide a necessary and sufficient stability criterion. However, when L is not normal, which is generic for the ODE system (1.1) obtained from semidiscrete PDE schemes, the eigenvalue analysis gives only necessary but possibly insufficient conditions for stability. This is due to the gap between the spectral radius and the operator norm. Therefore, the eigenvalue analysis may sometimes give misleading conclusions on the time step constraint [18, section 17.1, pp. 391–392] or the stability property [34].

To overcome the above-mentioned limitation, the energy method can be used as an alternative approach for stability analysis, which seeks certain energy identity or inequality. For implicit RK methods, their BN stability and algebraic stability ([4] and [39, Chapter IV.12]) were analyzed based on the energy method. For explicit RK methods, one stream of the research concerns the coercive operators [24],

which typically arise from diffusive problems such as method-of-lines schemes for the heat equation. It was shown that the Euler forward method is able to preserve the monotonous decay property $\left\|u^{n+1}\right\| \leq \|u^n\|$ under suitable time step constraint [13]. We will refer to this property as strong stability (sometimes also termed the monotonicity or monotonicity-preserving property in the literature [16, 20]). This stability property can be extended to all strong-stability-preserving RK methods [13, 21, 2, 19], which are constructed as convex combinations of Euler forward steps. In particular, those RK methods reducing to truncated Taylor expansions are all of such convex combination forms and thus strongly stable [13]. These arguments also coincide with the contractivity analysis under the so-called circle condition in [33, 23] and can be extended to nonlinear problems. However, such arguments may not be generally applied to noncoercive problems that commonly arise from semidiscrete schemes for wave type equations. A high-order energy expansion has to be carried out. Motivated by the studies on the third-order [37] and the fourth-order [34, 31] explicit RK methods, Sun and Shu [35] proposed a general framework on strong stability analysis for linear seminegative problems using the energy method. The essential idea of the novel framework [35] is to inductively apply a discrete analogue of integration by parts, which was inspired by the stability analysis of PDEs. In particular, it was proved in [35] that all linear RK methods corresponding to pth order truncated Taylor expansions are strongly stable if $p \equiv 3 \pmod 4$ and are not strongly stable if $p \equiv 1$ or $2 \pmod 4$. It is worth noting that the stability analysis in [35] is closely related to that of the RK discontinuous Galerkin schemes for linear advection equation by Xu et al. in [43, 42]. For nonlinear or nonautonomous problems, the requirement for strong stability may lead to order barriers [28, 29]. Remedy approaches to enforce strong stability were also studied recently, including the relaxation RK methods in [20, 32, 30] and the stabilization with artificial viscosity in [26, 36] and references therein.

It is worth particularly mentioning those implicit RK methods associated with the Padé approximations, which are the optimal rational approximation to the exponential function for given degrees of the numerator and denominator. The proof of A-stability of the diagonal Padé approximations may be dated back to [1]. Then it was shown that the first and the second subdiagonals in the Padé table are also A-stable [10, 11], but all the others are not A-stable [40]. It is also worth noting that some of the Padé approximations correspond to the stability functions of certain collocation methods such as the Gauss methods, the Radau methods, and the Lobatto methods [39, Table 5.13, p. 82]. The analysis of algebraic stability for those collocation methods [15] could also lead to the L^2-stability of the corresponding Padé approximations.

In this paper, we generalize the stability analysis of *explicit* RK methods in [35] and establish a systematic theoretical framework for analyzing *general implicit and explicit* RK methods. The efforts and novelty of this paper are summarized as follows.

- We present a universal framework to derive the energy identity of a generic RK method for general linear seminegative systems (1.1). The energy identity provides a precise characterization on how the energy law (1.4) is approximated and whether the energy dissipation property is preserved in the RK discretization. As a result, the established energy identities lead to weak and strong stability criteria of RK methods.
- Our framework is motivated by a series expansion of the *continuous* energy law (1.4) and a *discrete* analogue of *integration by parts* technique. Hence we also refer to our energy identities as *discrete energy laws*. Our analyses in some special cases establish a close connection between the continuous and discrete energy laws. The findings clearly demonstrate the unity of continuous and discrete objects.

- Besides the different motivations, some other aspects of our framework are also quite different from those of the eigenvalue analysis and the traditional energy approaches such as the algebraic stability analysis. In our discrete energy laws, the energy dissipation is carefully expanded in terms of the proposed seminorm $[\![\cdot]\!]$ associated with the operator L. Moreover, our expansion is formulated as a high-order polynomial of the time stepsize τ, which can be compared with the infinite series expansion in the continuous case.
- Moreover, we discover the unified discrete energy law for all the diagonal Padé approximations of arbitrary orders. Such a unified energy law is established based on an analytical Cholesky type decomposition of a class of symmetric matrices. The structure of the matrices is extremely complicated and their elements involve complex summations of factorial products; see (5.3). As a result, the discovery of the unified energy law and the proof of the decomposition are highly nontrivial and challenging; see Theorem 5.1 and its proof in subsection 5.3. Besides, our analyses involve the construction of technical combinatorial identities and some novel techniques from the theory of hypergeometric series, which seem to be rarely used in previous RK stability analyses and may shed new light on future developments in this direction.
- It is worth noting that the proposed framework applies to a generic RK method, which can be either implicit or explicit, unconditionally stable (A-stable), or conditionally stable (not A-stable). We provide several specific examples of implicit methods in section 4 to further understand the proposed discrete energy laws. A few numerical examples are also given in section 6 to confirm the theoretical results.

The paper is organized as follows. We study the continuous energy law in section 2 and present the systematic theoretical framework in section 3 to derive the discrete energy laws of general RK methods and the stability analysis. Examples on implicit RK methods are given in section 4. We derive the unified discrete energy law of diagonal Padé approximations in section 5 and present numerical results in section 6 before conclusions in section 7. For better readability, some technical proofs are presented in the appendices.

2. Energy law at continuous level. In this section, we derive a series expansion of the continuous energy law (1.4) for the linear seminegative system (1.1). The main result is given below.

THEOREM 2.1. *The energy law of the linear seminegative problem* (1.1) *has the series expansion*

$$\|u(t^n+\tau)\|^2 - \|u(t^n)\|^2 = -\sum_{k=0}^{\infty} \widehat{d}_k \tau^{2k+1} \left[\!\!\left[L^k \widehat{u}^{(k)} \right]\!\!\right]^2, \tag{2.1}$$

where

$$\widehat{u}^{(k)} = \sum_{j=k}^{\infty} \widehat{\mu}_{k,j} (\tau L)^{j-k} u(t^n) \tag{2.2}$$

with $\widehat{d}_k$ *and* $\widehat{\mu}_{k,j}$ *defined by*

$$\widehat{d}_k := \frac{(k!)^2}{(2k)!(2k+1)!}, \qquad \widehat{\mu}_{k,j} := \frac{(2k+1)!j!}{k!(j-k)!(k+j+1)!} \qquad \forall k, j \in \mathbb{N},\ j \geq k. \tag{2.3}$$

The significance of the expansion (2.1) lies in that each term in the expansion clearly shows the energy dissipation order with respect to τ. This will help to gain

some insights on deriving similar expansions for the discrete energy laws of RK methods in section 3. Theorem 2.1 will also be useful for establishing a connection between the continuous energy law and the discrete energy laws in subsection 5.2. It is also worth noting that the infinite series $\widehat{u}^{(k)}$ in (2.2) is well-defined, because

$$\left\|\widehat{u}^{(k)}\right\| \le \frac{(2k+1)!}{k!}\sum_{j=k}^{\infty}\frac{(\tau\|L\|)^{j-k}}{(j-k)!}\|u(t^n)\| = \frac{(2k+1)!}{k!}e^{\tau\|L\|}\|u(t^n)\| < \infty,$$

where and hereafter the operator norm of L is defined as $\|L\| := \sup\{\|Lv\| : \|v\| \le 1, v \in V\}$.

The proof of Theorem 2.1 is fairly technical and relies on Lemmas 2.2 and B.1. To improve the readability of the paper, we place the detailed proof of Theorem 2.1 in Appendix C, right after the proofs of Lemmas 2.2 and B.1 in Appendices A and B, respectively.

LEMMA 2.2. *Let N be a nonnegative integer. Assume that the matrix $\boldsymbol{\Upsilon} = (\gamma_{i,j})_{i,j=0}^{N}$ is negative semidefinite with the Cholesky type decomposition $\boldsymbol{\Upsilon} = -\mathbf{U}^{\top}\mathbf{D}\mathbf{U}$, where $\mathbf{U} = (\mu_{k,j})_{k,j=0}^{N}$ is an upper triangular matrix and $\mathbf{D} = \mathrm{diag}(\{d_k\}_{k=0}^{N})$ is a diagonal matrix with nonnegative entries. Then for any $v \in V$, it holds that*

$$\sum_{i=0}^{N}\sum_{j=0}^{N}\gamma_{i,j}\tau^{i+j+1}\left[L^i v, L^j v\right] = -\sum_{k=0}^{N} d_k \tau^{2k+1}\left[\!\left[L^k v^{(k)}\right]\!\right]^2 \le 0, \tag{2.4}$$

where $v^{(k)} = \sum_{j=k}^{N}\mu_{k,j}(\tau L)^{j-k}v$.

Notice that Lemma 2.2 is a universal result applicable to any negative semidefinite matrices $\boldsymbol{\Upsilon}$. Lemma 2.2 is not only used for a special matrix $\widehat{\gamma}_{i,j} = -\frac{1}{i!j!(i+j+1)}$ in the proof of Theorem 2.1, but will also be used for the general matrix $\gamma_{i,j}$ defined in (3.10) to derive the discrete energy laws of RK methods in Theorem 3.4.

Remark 2.3. The energy decay property $\|u(t^n+\tau)\|^2 - \|u(t^n)\|^2 = \left\|e^{\tau L}u(t^n)\right\|^2 - \|u(t^n)\|^2 \le 0$ can be equivalently expressed as $(e^{\tau L})^{\top}e^{\tau L} - I \le O$ is negative semidefinite. Theorem 2.1 gives a more precise characterization of this property by expanding it into an infinite series of negative semidefinite operators

$$(e^{\tau L})^{\top}e^{\tau L} - I = \sum_{k=0}^{\infty}\widehat{d}_k\tau^{2k+1}\widehat{U}_k^{\top}(L^{\top}+L)\widehat{U}_k \le O \quad \text{with} \quad \widehat{U}_k := L^k\sum_{j=k}^{\infty}\widehat{\mu}_{k,j}(\tau L)^{j-k}, \tag{2.5}$$

where $\widehat{d}_k$ and $\widehat{\mu}_{k,j}$ are defined in (2.3) and $L^{\top}$ is the adjoint operator of L. The identity (2.5) directly follows from (2.1) in Theorem 2.1, by noting that $u(t^n)$ can be arbitrarily taken in the space V.

3. Discrete energy laws and stability of Runge–Kutta methods. We consider the RK discretizations to the seminegative system (1.1). Our goal is to establish a unified framework for deriving the discrete energy laws satisfied by the numerical solutions of the RK methods. The discrete energy laws are analogues of the continuous energy law (2.1) and will be very useful for understanding and analyzing the stability of RK methods.

In general, an RK method for the linear autonomous system (1.1) can be formulated as

$$u^{n+1} = \mathcal{R}(\tau L)u^n, \tag{3.1}$$

where u^n denotes the numerical solution at the nth time level $t = t^n$, and $\tau = t^{n+1} - t^n$ is the time stepsize. Here $\mathcal{R}(Z)$ is the stability function corresponding to a rational approximation of e^Z given by

$$\mathcal{R}(Z) = (\mathcal{Q}(Z))^{-1}\mathcal{P}(Z) \tag{3.2}$$

with $\mathcal{P}(Z)$ and $\mathcal{Q}(Z)$ being s_pth and s_qth order polynomials of Z, namely,

$$\mathcal{P}(Z) = \sum_{i=0}^{s} \theta_i Z^i \qquad \text{with } \theta_i = 0 \text{ for } i > s_p, \tag{3.3a}$$

$$\mathcal{Q}(Z) = \sum_{i=0}^{s} \vartheta_i Z^i \qquad \text{with } \vartheta_i = 0 \text{ for } i > s_q, \tag{3.3b}$$

where $s := \max\{s_p, s_q\}$, and a normalization is typically used such that $\theta_0 = \vartheta_0 = 1$. For convenience, we denote $P := \mathcal{P}(\tau L)$ and $Q := \mathcal{Q}(\tau L)$. Note that the operators L, P, and Q^{-1} commute with each other.

Remark 3.1. In the special case that $s_q = 0$, namely, $\mathcal{R}(Z)$ is a polynomial approximation of e^Z, then $Q = I$ is the identity operator, and the scheme (3.1) is an explicit RK method, whose stability was studied in [35] via the energy approach. When $s_q \geq 1$, the RK method (3.1) is implicit, which is the particular focus of the present paper.

In our following analysis of stability and energy laws, we always assume that the equations of the given (implicit) RK method are uniquely solvable, namely, the operator Q is invertible and hence (3.1) is well-defined. This is a reasonable and basic assumption before one starts to consider the stability of the RK method. Certainly, such unique solvability is an important topic and may require an additional condition on τ; we will give some discussions on this topic in subsection 3.3.

3.1. Discrete energy laws. We first give a lemma on the energy change of the RK method (3.1).

LEMMA 3.2. *The solution of the RK method* (3.1) *satisfies the following identity:*

$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 = \sum_{i=0}^{s}\sum_{j=0}^{s} \alpha_{i,j}\tau^{i+j} \left\langle L^i w^n, L^j w^n \right\rangle, \tag{3.4}$$

where $w^n := Q^{-1}u^n$ *and* $\alpha_{i,j} := \theta_i\theta_j - \vartheta_i\vartheta_j$.

Proof. Some simple algebraic manipulations give

$$\begin{aligned} \left\|u^{n+1}\right\|^2 &= \left\|Q^{-1}Pu^n\right\|^2 = \left\|PQ^{-1}u^n\right\|^2 = \|u^n\|^2 + \left\|PQ^{-1}u^n\right\|^2 - \left\|QQ^{-1}u^n\right\|^2 \\ &= \|u^n\|^2 + \|Pw^n\| - \|Qw^n\|^2 . \end{aligned} \tag{3.5}$$

Note that

$$\|Pw^n\|^2 = \left\langle \sum_{i=0}^{s} \theta_i(\tau L)^i w^n, \sum_{j=0}^{s} \theta_j(\tau L)^j w^n \right\rangle = \sum_{i=0}^{s}\sum_{j=0}^{s} \theta_i\theta_j\tau^{i+j} \left\langle L^i w^n, L^j w^n \right\rangle,$$

and similarly $\|Qw^n\|^2 = \sum_{i=0}^{s}\sum_{j=0}^{s} \vartheta_i\vartheta_j\tau^{i+j} \left\langle L^i w^n, L^j w^n \right\rangle$. Substituting these expansions into (3.5) gives (3.4) and completes the proof. ∎

However, from the energy identity (3.4), it is very difficult to judge whether the energy $\|u^n\|^2$ always decays or not, because the sign of each term $\langle L^i w^n, L^j w^n\rangle$ in (3.4) is unclear and indeterminate. In order to address this difficulty, we would like to reformulate $\langle L^i w^n, L^j w^n\rangle$ into a linear combination of some terms of form $\|L^k w^n\|^2$ and $[L^k w^n, L^l w^n]$. Such a reformulation procedure can be completed by repeatedly using a discrete analogue of the integration by parts formula

$$\langle w, Lv\rangle = -\langle Lw, v\rangle - [w, v], \tag{3.6}$$

which follows from the definition (1.2) and gives

$$\langle L^i w^n, L^j w^n\rangle = \begin{cases} \|L^i w^n\|^2 & \text{if } j = i, \\ -\frac{1}{2}\left[\!\left[L^i w^n\right]\!\right]^2 & \text{if } j = i+1, \\ -\langle L^{i+1} w^n, L^{j-1} w^n\rangle - [L^i w^n, L^{j-1} w^n] & \text{otherwise.} \end{cases} \tag{3.7}$$

See [35, Proposition 2.1] for a proof of (3.7). It is worth noting that such a discrete version of integration by parts is inspired by approximating the spatial derivative ∂_x with L.

Recursively applying (3.7) to reformulate the terms $\langle L^i v, L^j v\rangle$ in (3.4), we obtain an energy identity in the following form.

LEMMA 3.3. *For the solution of the RK method* (3.1)*, the following identity holds:*

$$\|u^{n+1}\|^2 - \|u^n\|^2 = \sum_{k=0}^{s} \beta_k \tau^{2k} \|L^k w^n\|^2 + \sum_{i=0}^{s-1}\sum_{j=0}^{s-1} \gamma_{i,j} \tau^{i+j+1} [L^i w^n, L^j w^n], \tag{3.8}$$

where β_k and $\gamma_{i,j}$ are computed from the values of $\alpha_{i,j} = \theta_i\theta_j - \vartheta_i\vartheta_j$ via the formulae

$$\beta_k = \sum_{\ell=\max\{0,2k-s\}}^{\min\{2k,s\}} \alpha_{\ell,2k-\ell}(-1)^{k-\ell}, \tag{3.9}$$

$$\gamma_{i,j} = \sum_{\ell=\max\{0,i+j+1-s\}}^{\min\{i,j\}} (-1)^{\min\{i,j\}+1-\ell} \alpha_{\ell,i+j+1-\ell}. \tag{3.10}$$

We remark that for a given RK method, $\{\theta_i\}$ and $\{\vartheta_i\}$ are given, and $\{\beta_k\}$ and $\{\gamma_{i,j}\}$ are determined by (3.9)–(3.10). A constructive proof of Lemma 3.3 can be found in [12]. In the following, we present a different and shorter proof via direct verification.

Proof. Using (3.6), we can rewrite the right-hand side of (3.8) as

$$\begin{aligned} &\sum_{k=0}^{s} \beta_k \tau^{2k} \|L^k w^n\|^2 - \sum_{i=0}^{s-1}\sum_{j=0}^{s-1} \gamma_{i,j} \tau^{i+j+1} \langle L^i w^n, L^{j+1} w^n\rangle \\ &\quad - \sum_{i=0}^{s-1}\sum_{j=0}^{s-1} \gamma_{i,j} \tau^{i+j+1} \langle L^{i+1} w^n, L^j w^n\rangle \\ &= \sum_{k=0}^{s} \beta_k \tau^{2k} \langle L^k w^n, L^k w^n\rangle - \sum_{i=0}^{s-1}\sum_{j=1}^{s} \gamma_{i,j-1} \tau^{i+j} \langle L^i w^n, L^j w^n\rangle \\ &\quad - \sum_{i=1}^{s}\sum_{j=0}^{s-1} \gamma_{i-1,j} \tau^{i+j} \langle L^i w^n, L^j w^n\rangle. \end{aligned}$$

By comparing the expansion coefficients with (3.4) in Lemma 3.2, it suffices to verify the identity

$$\alpha_{i,j} = \beta_i 1_{\{i=j\}} - \gamma_{i,j-1} 1_{\{i<s,j>0\}} - \gamma_{i-1,j} 1_{\{i>0,j<s\}}, \qquad 0 \le i,j \le s, \tag{3.11}$$

where $1_{\{\cdot\}}$ is the indicator function. Since $\alpha_{i,j} = \alpha_{j,i}$ and $\gamma_{i,j} = \gamma_{j,i}$, both sides of (3.11) are symmetric with (i,j). Hence we only need to verify (3.11) for $0 \le i \le j \le s$ by verifying the following two cases:

- Case 1: $0 \le i = j \le s$. In this case, after applying (3.9)–(3.10), one can show that the right-hand side of (3.11) becomes

$$\begin{aligned}
&\beta_i - 2\gamma_{i-1,i} 1_{\{0<i<s\}} \\
&\quad = \sum_{\ell=\max\{0,2i-s\}}^{\min\{2i,s\}} (-1)^{i-\ell} \alpha_{\ell,2i-\ell} - 1_{\{0<i<s\}} \left(2 \sum_{\ell=\max\{0,2i-s\}}^{i-1} (-1)^{i-\ell} \alpha_{\ell,2i-\ell} \right) \\
&\quad = \sum_{\ell=\max\{0,2i-s\}}^{\min\{2i,s\}} (-1)^{i-\ell} \alpha_{\ell,2i-\ell} - 1_{\{0<i<s\}} \\
&\qquad \times \left(\sum_{\ell=\max\{0,2i-s\}}^{i-1} (-1)^{i-\ell} \alpha_{\ell,2i-\ell} + \sum_{k=i+1}^{\min\{2i,s\}} (-1)^{k-i} \alpha_{2i-k,k} \right),
\end{aligned}$$

  which is equal to $\alpha_{i,i}$ after checking the three cases: $0 < i < s$, $i = 0$, and $i = s$, respectively.
- Case 2: $0 \le i < j \le s$. In this case, $1_{\{i<s,j>0\}} = 1$ and $i \le j - 1$. Together with (3.10), the right-hand side of (3.11) then reduces to

$$\begin{aligned}
&-\gamma_{i,j-1} - \gamma_{i-1,j} 1_{\{i>0,j<s\}} \\
&\quad = \sum_{\ell=\max\{0,i+j-s\}}^{i} (-1)^{i-\ell} \alpha_{\ell,i+j-\ell} - 1_{\{i>0,j<s\}} \sum_{\ell=\max\{0,i+j-s\}}^{i-1} (-1)^{i-\ell} \alpha_{\ell,i+j-\ell},
\end{aligned}$$

  which is equal to $\alpha_{i,j}$ by checking the three cases: $0 < i < j < s$, $i = 0$, and $j = s$, respectively.

In summary, the identity (3.11) holds, and (3.8) is equivalent to (3.4). The proof is completed. ∎

Note that the first term at the right-hand side of (3.8) has a similar format as that in the continuous energy law (2.1). Next, we would like to reformulate the last term of (3.8) by using Lemma 2.2. Define

$$\mathbf{B} := \operatorname{diag}(\{\beta_k\}_{k=0}^{s}) \quad \text{and} \quad \boldsymbol{\Upsilon} := (\gamma_{i,j})_{i,j=0}^{s-1}$$

with β_k and $\gamma_{i,j}$ given by (3.9)–(3.10), respectively. However, for some RK methods the symmetric matrix $\boldsymbol{\Upsilon}$ is not necessarily negative semidefinite, so that its Cholesky type decomposition required in Lemma 2.2 may not exist. In case this happens, one can overcome such a problem by subtracting a diagonal matrix. We finally obtain the following practical discrete energy law (3.12) for general RK methods.

THEOREM 3.4 (energy identity). *Assume that $\tilde{\boldsymbol{\Upsilon}} = \boldsymbol{\Upsilon} - \boldsymbol{\Delta}$ is negative semidefinite for some diagonal matrix $\boldsymbol{\Delta} = \operatorname{diag}(\{\delta_k\}_{k=0}^{s-1})$ with $\delta_k \ge 0$ for $0 \le k \le s-1$, so that the symmetric matrix $\tilde{\boldsymbol{\Upsilon}}$ admits the Cholesky type decomposition $\tilde{\boldsymbol{\Upsilon}} = -\tilde{\mathbf{U}}^\top \tilde{\mathbf{D}} \tilde{\mathbf{U}}$, where*

$\tilde{\mathbf{U}} = (\tilde{\mu}_{k,i})_{k,i=0}^{s-1}$ is an upper triangular matrix with $\mu_{k,k} = 1$ and $\tilde{\mathbf{D}} = \mathrm{diag}(\{\tilde{d}_k\}_{k=0}^{s-1})$ with $\tilde{d}_k \geq 0$ for $0 \leq k \leq s-1$. The solution of the RK method (3.1) *satisfies the following energy identity:*

(3.12)
$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 = \sum_{k=0}^{s} \beta_k \tau^{2k} \left\|L^k w^n\right\|^2 - \sum_{k=0}^{s-1} \tilde{d}_k \tau^{2k+1} \left[\!\left[L^k u^{(k)}\right]\!\right]^2 + \sum_{k=0}^{s-1} \delta_k \tau^{2k+1} \left[\!\left[L^k w^n\right]\!\right]^2,$$

where $u^{(k)} := \sum_{j=k}^{s} \tilde{\mu}_{k,j} (\tau L)^{j-k} w^n = \sum_{j=k}^{s} \tilde{\mu}_{k,j} (\tau L)^{j-k} Q^{-1} u^n$.

Proof. Denote $\tilde{\boldsymbol{\Upsilon}} =: (\tilde{\gamma}_{i,j})_{i,j=0}^{s-1}$. Then it follows from (3.8) and $\boldsymbol{\Upsilon} = \tilde{\boldsymbol{\Upsilon}} + \boldsymbol{\Delta}$ that

$$\begin{aligned}
\left\|u^{n+1}\right\|^2 - \|u^n\|^2 &= \sum_{k=0}^{s} \beta_k \tau^{2k} \left\|L^k w^n\right\|^2 + \sum_{i=0}^{s-1}\sum_{j=0}^{s-1} \gamma_{i,j} \tau^{i+j+1} \left[L^i w^n, L^j w^n\right] \\
&= \sum_{k=0}^{s} \beta_k \tau^{2k} \left\|L^k w^n\right\|^2 + \sum_{i=0}^{s-1}\sum_{j=0}^{s-1} \tilde{\gamma}_{i,j} \tau^{i+j+1} \left[L^i w^n, L^j w^n\right] \\
&\quad + \sum_{k=0}^{s-1} \delta_k \tau^{2k+1} \left[\!\left[L^k w^n\right]\!\right]^2.
\end{aligned}$$

Using Lemma 2.2 to reformulate the second term yields (3.12). ∎

Examples of the discrete energy law (3.12) for several specific RK schemes will be given in section 4.

3.2. Stability analysis. This subsection applies the discrete energy law (3.12) in Theorem 3.4 to analyze the stability of RK methods.

First, consider a special case: both $\boldsymbol{\Upsilon}$ and $\mathbf{B}$ are negative semidefinite. We obtain the unconditional strong stability of the corresponding RK method from the discrete energy law (3.12).

THEOREM 3.5 (unconditional strong stability). *If the RK method* (3.1) *satisfies that $\boldsymbol{\Upsilon}$ and $\mathbf{B}$ are both negative semidefinite, then the RK method* (3.1) *is unconditionally strongly stable, namely,*

$$\left\|u^{n+1}\right\|^2 \leq \|u^n\|^2 \qquad \forall \tau \geq 0. \tag{3.13}$$

Proof. When $\boldsymbol{\Upsilon}$ is negative semidefinite, Theorem 3.4 holds with $\boldsymbol{\Delta} = \mathbf{O}$, namely, we can take $\delta_k = 0$, so that the energy identity (3.12) becomes

$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 = \sum_{k=0}^{s} \beta_k \tau^{2k} \left\|L^k w^n\right\|^2 - \sum_{k=0}^{s-1} \tilde{d}_k \tau^{2k+1} \left[\!\left[L^k u^{(k)}\right]\!\right]^2.$$

This yields (3.13), because $\tilde{d}_k \geq 0$ and $\beta_k \leq 0$ as $\mathbf{B}$ is negative semidefinite. ∎

In the rest of this section, we discuss the general case that $\boldsymbol{\Upsilon}$ is not necessarily negative semidefinite, and we shall use the energy law (3.12) to derive several stability criteria under some constraint on the time stepsize τ. For simplicity, we will denote $\tau \|L\| =: \lambda$ and use the notation λ_0 and C to represent generic positive constants, which are independent of τ and $\|L\|$ but may depend on θ_i, ϑ_i, and s. The values of λ_0 and C may vary at different places.

LEMMA 3.6 (energy estimate). *Let ζ be the index of the first nonzero element in $\{\beta_k\}_{k=0}^s$. Let ρ be the largest index such that the ρth order principle submatrix $(\gamma_{i,j})_{i,j=0}^{\rho-1}$ is negative semidefinite. There exists a positive constant c_ρ such that*

(3.14)
$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 \le \left(\beta_\zeta + \lambda^2 g_\beta(\lambda)\right)\tau^{2\zeta}\left\|L^\zeta w^n\right\|^2 + \lambda c_\rho(1+\lambda^2 g_\rho(\lambda))\tau^{2\rho}\left\|L^\rho w^n\right\|^2,$$

where $g_\beta(\lambda) := \sum_{i=0}^{s-\zeta-1}\beta_{i+\zeta+1}\lambda^{2i}$ and $g_\rho(\lambda) := \sum_{i=0}^{s-\rho-2}\lambda^{2i}$ are polynomials of $\lambda := \tau\|L\|$.

Proof. Since the ρth order principle submatrix of $\boldsymbol{\Upsilon}$ is negative semidefinte, there exists a positive constant c_ρ such that the symmetric matrix

$$\tilde{\boldsymbol{\Upsilon}} := \boldsymbol{\Upsilon} - \frac{1}{2}\operatorname{diag}\{\underbrace{0,\dots,0}_{\rho},\underbrace{c_\rho,\dots c_\rho}_{s-\rho}\} =: \boldsymbol{\Upsilon} - \boldsymbol{\Delta}$$

is negative semidefinite. According to the energy law (3.12) in Theorem 3.4, we have

(3.15)
$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 = \sum_{k=0}^{s}\beta_k\tau^{2k}\left\|L^k w^n\right\|^2 - \sum_{k=0}^{s-1}\tilde{d}_k\tau^{2k+1}\left[\!\left[L^k u^{(k)}\right]\!\right]^2 + \frac{c_\rho}{2}\sum_{k=\rho}^{s-1}\tau^{2k+1}\left[\!\left[L^k w^n\right]\!\right]^2.$$

For the first term on the right-hand side of (3.15), using the Cauchy–Schwarz inequality gives

$$\begin{aligned}\sum_{k=0}^{s}\beta_k\tau^{2k}\left\|L^k w^n\right\|^2 &= \sum_{k=\zeta}^{s}\beta_k\left\|(\tau L)^k w^n\right\|^2 \le \sum_{k=\zeta}^{s}\beta_k\left(\tau\|L\|\right)^{2(k-\zeta)}\tau^{2\zeta}\left\|L^\zeta w^n\right\|^2 \\ &= \left(\beta_\zeta + \lambda^2\sum_{i=0}^{s-\zeta-1}\beta_{i+\zeta+1}\lambda^{2i}\right)\tau^{2\zeta}\left\|L^\zeta w^n\right\|^2.\end{aligned} \tag{3.16}$$

For the second term, since $\tilde{d}_k \ge 0$ for $0 \le k \le s-1$, we have

$$-\sum_{k=0}^{s-1}\tilde{d}_k\tau^{2k+1}\left[\!\left[L^k u^{(k)}\right]\!\right]^2 \le 0. \tag{3.17}$$

For the last term, one can again utilize the Cauchy–Schwarz inequality to obtain

$$\begin{aligned}\frac{c_\rho}{2}\sum_{k=\rho}^{s-1}\tau^{2k+1}\left[\!\left[L^k w^n\right]\!\right]^2 &\le c_\rho\sum_{k=\rho}^{s-1}\left(\tau\|L\|\right)^{2(k-\rho)+1}\|(\tau L)^\rho w^n\|^2 \\ &= \lambda c_\rho\left(1+\lambda^2\sum_{i=0}^{s-\rho-2}\lambda^{2i}\right)\tau^{2\rho}\left\|L^\rho u^n\right\|^2.\end{aligned} \tag{3.18}$$

Combining the estimates in (3.16)–(3.18) with (3.15) gives (3.14) and completes the proof. ∎

THEOREM 3.7 (conditional stability criteria). *Denote $\lambda := \tau\|L\|$. Let ζ and ρ be the indexes defined in Lemma 3.6 and $\kappa := \min\{2\zeta, 2\rho+1\}$. We have the following stability criteria for a generic RK method:*

1. *The RK method* (3.1) *is weakly*(κ) *stable, namely,* $\left\|u^{n+1}\right\|^2 \le (1+C\lambda^\kappa)\left\|u^n\right\|^2$, *under a time step constraint* $\lambda \le \lambda_0$ *for some positive constant* λ_0. *Furthermore, if* λ^κ/τ *is bounded, or equivalently,* $\tau\left\|L\right\|^{1+1/(\kappa-1)} \le \lambda_0$ *for some positive constant* λ_0, *then* $\left\|u^n\right\|^2 \le e^{Ct^n}\left\|u^0\right\|^2$, *where* $t^n = n\tau$.
2. *If* $\zeta \le \rho$ *and* $\beta_\zeta < 0$, *then the RK method* (3.1) *is strongly stable, namely,* $\left\|u^{n+1}\right\|^2 \le \left\|u^n\right\|^2$, *under a time step constraint* $\lambda \le \lambda_0$ *for some positive constant* λ_0.
3. *If* $\beta_\zeta > 0$, *then the RK method* (3.1) *is not strongly stable for a generic seminegative system* (1.1), *namely, there exist a linear seminegative operator* L *and a positive constant* λ_0 *such that* $\left\|\mathcal{R}(\tau L)\right\| > 1$ *for any* $\lambda \in (0, \lambda_0]$.

Proof. For the first part on the weak(κ) stability, we observe that

$$\left\|u^n\right\| = \left\|Qw^n\right\| = \left\|w^n + \sum_{k=1}^{s}\vartheta_k(\tau L)^k w^n\right\| \ge \left(1 - \sum_{k=1}^{s}|\vartheta_k|(\tau\left\|L\right\|)^k\right)\left\|w^n\right\|.$$

When $\tau\left\|L\right\| = \lambda$ is sufficiently small, we have $\left\|u^n\right\| \ge \frac{1}{2}\left\|w^n\right\|$ and $\left\|w^n\right\| \le 2\left\|u^n\right\|$. It follows that $\tau^{2k}\left\|L^k w^n\right\|^2 \le \lambda^{2k}\left\|w^n\right\|^2 \le 4\lambda^{2k}\left\|u^n\right\|^2$. Similar arguments yield $\beta_\zeta + \lambda^2 g_\beta(\lambda) \le 2|\beta_\zeta|$ and $c_\rho(1+\lambda^2 g_\rho(\lambda)) \le 2c_\rho$ when λ is sufficiently small. These together with the energy estimate in Lemma 3.6 imply

$$\begin{aligned}\left\|u^{n+1}\right\|^2 &\le \left\|u^n\right\|^2 + 2|\beta_\zeta|\tau^{2\zeta}\left\|L^\zeta w^n\right\|^2 + 2c_\rho\lambda\tau^{2\rho}\left\|L^\rho w^n\right\|^2\\ &\le \left(1 + 8|\beta_\zeta|\lambda^{2\zeta} + 8c_\rho\lambda^{2\rho+1}\right)\left\|u^n\right\|^2 \le (1+C\lambda^\kappa)\left\|u^n\right\|^2\end{aligned}$$

under the constraint $\lambda \le \lambda_0$ for some positive constant λ_0. Furthermore, if λ^κ/τ is bounded, we have

$$\left\|u^n\right\|^2 \le (1+C\lambda^\kappa)^n\left\|u^0\right\|^2 = (1+C\lambda^\kappa)^{\lambda^{-\kappa}\cdot t_n\cdot\frac{\lambda^\kappa}{\tau}}\left\|u^0\right\|^2 \le e^{Ct^n\lambda^\kappa/\tau}\left\|u^0\right\|^2 \le e^{Ct^n}\left\|u^0\right\|^2.$$

We then turn to prove the second part of the theorem. Observe that $\lambda g_\beta(\lambda) \le 1$ and $\lambda^2 g_\rho(\lambda) \le 1$ when $\lambda \le \widehat{\lambda}_0$ for some constant $\widehat{\lambda}_0$. Thanks to Lemma 3.6, when $\zeta \le \rho$ and $\lambda \le \widehat{\lambda}_0$ we then have

$$\begin{aligned}\left\|u^{n+1}\right\|^2 - \left\|u^n\right\|^2 &\le (\beta_\zeta + \lambda)\tau^{2\zeta}\left\|L^\zeta w^n\right\|^2 + 2c_\rho\lambda\tau^{2\rho}\left\|L^\rho w^n\right\|^2\\ &\le (\beta_\zeta + \lambda)\tau^{2\zeta}\left\|L^\zeta w^n\right\|^2 + 2c_\rho\lambda\tau^{2\rho}\left\|L\right\|^{2(\rho-\zeta)}\left\|L^\zeta w^n\right\|^2\\ &\le \left(\beta_\zeta + \lambda + 2\lambda c_\rho\widehat{\lambda}_0^{2(\rho-\zeta)}\right)\tau^{2\zeta}\left\|L^\zeta w^n\right\|^2,\end{aligned}$$

where the last term is nonpositive if $\lambda \le |\beta_\zeta|/(1 + 2c_\rho\widehat{\lambda}_0^{2(\rho-\zeta)})$. We therefore obtain $\left\|u^{n+1}\right\|^2 \le \left\|u^n\right\|^2$ under the constraint $\lambda \le \lambda_0$ with $\lambda_0 := \min\{\widehat{\lambda}_0, |\beta_\zeta|/(1 + 2c_\rho\widehat{\lambda}_0^{2(\rho-\zeta)})\}$.

For the third part, one can consider a special operator L satisfying $L^\zeta Q^{-1} \ne O$ but $L^\top + L = O$, so that the last term in (3.8) vanishes. It then follows from Lemma 3.3 that

$$\left\|\mathcal{R}(\tau L)u^n\right\|^2 - \left\|u^n\right\|^2 = \sum_{k=\zeta}^{s}\beta_k\tau^{2k}\left\|L^k Q^{-1}u^n\right\|^2 \ge \left(\beta_\zeta - \sum_{k=\zeta+1}^{s}|\beta_k|\lambda^{2k}\right)\tau^{2\zeta}\left\|L^\zeta Q^{-1}u^n\right\|^2.$$

Hence when λ is sufficiently small, we have $\left\|\mathcal{R}(\tau L)u^n\right\| / \left\|u^n\right\| > 1$ for all u^n satisfying $L^\zeta Q^{-1}u^n \ne 0$, which implies $\left\|\mathcal{R}(\tau L)\right\| > 1$. The proof is completed. ∎

Remark 3.8. If the system (1.1) is obtained from spatially semidiscrete schemes for linear hyperbolic conservation laws, then we have $\|L\| = \mathcal{O}(h^{-1})$, where h is the spatial mesh size. In this case, the time step constraint $\lambda = \tau\|L\| \le \lambda_0$ in Theorem 3.7 becomes the Courant–Friedrichs–Lewy (CFL) condition $\tau \le Ch$. The time step constraint for weak(κ) stability, $\tau\|L\|^{1+1/(\kappa-1)} \le \lambda_0$, becomes $\tau \le Ch^{1+1/(\kappa-1)}$.

Remark 3.9. The weak stability condition $\tau\|L\|^{1+1/(\kappa-1)} \le \lambda_0$ in Theorem 3.7 is also necessary, in the sense that the power of $\|L\|$ cannot be improved in general. To see this, we consider the Euler forward method $u^{n+1} = (I + \tau L)u^n$ with an antisymmetric matrix L. For this method, $\kappa = 2$ and $1 + 1/(\kappa - 1) = 2$. Let ρ_{radi} denote the spectral radius. Because $L^\top + L = O$, one has

$$\begin{aligned}\|I + \tau L\|^2 =& \rho_{\text{radi}}\left((I + \tau L)^\top (I + \tau L)\right) = \rho_{\text{radi}}\left(I + \tau(L^\top + L) + \tau^2 L^\top L\right)\\ =& \rho_{\text{radi}}(I + \tau^2 L^\top L) = 1 + \tau^2 \rho_{\text{radi}}(L^\top L) = 1 + \tau^2\|L\|^2.\end{aligned} \tag{3.19}$$

Let us fix the final time $T = 1$ and denote $\tau = T/n = 1/n$. Since $u^n = (I + \tau L)^n u^0$, to ensure the weak stability for an arbitrary u^0, we require the time step constraint such that $\|(I + \tau L)^n\|$ is bounded. Let $\tau\|L\|^m = \lambda_0$ for a fixed constant λ_0. Note that $A := I + \tau L$ is normal and $\|A^n\| = \|A\|^n$ for any normal matrices. Then, under the time step condition $\tau\|L\|^m = \lambda_0$, we have

$$\begin{aligned}\|(I + \tau L)^n\|^2 = \left(\|I + \tau L\|^2\right)^n \overset{(3.19)}{=} \left(1 + \tau^2\|L\|^2\right)^n = \left(1 + \lambda_0^{2/m}\tau^{2-2/m}\right)^n\\ = \left(1 + \lambda_0^{2/m} n^{2/m-2}\right)^n.\end{aligned}$$

As $n \to \infty$ (i.e., $\tau \to 0^+$), the above term is bounded if and only if $2/m - 2 \le -1$, namely, $m \ge 2$. The system (1.1) with an antisymmetric L may arise from semidiscrete schemes, e.g., the central difference scheme $\frac{\mathrm{d}}{\mathrm{d}t}u_j = \frac{1}{2h}(u_{j+1} - u_{j-1})$, the Fourier spectral method, or the discontinuous Galerkin method with the central flux, for the linear convection equation $\partial_t u = \partial_x u$ with periodic boundary conditions. In this case, $\|L\| = \mathcal{O}(h^{-1})$, and according to the above analysis, the Euler forward method is stable under the condition $\tau \le Ch^m$ with $m \ge 2$ and is unstable if $\tau = Ch^m$ with $m < 2$.

Although the weak stability condition in Theorem 3.7 is considered to be sharp for general L, it is possible to improve this condition for specific problems in which L admits special structures. For example, in [43, section 5.2], it was pointed out that the stability estimates can be improved in the context of RK discontinuous Galerkin methods for linear advection if L is constructed with low-order spatial elements.

Remark 3.10. The stability analyses in Theorem 3.7 and [35] are closely connected with the L^2-stability analysis of RK discontinuous Galerkin schemes for the linear advection equation by Xu and co-authors in [43, 42], where the weak(κ) stability was systematically studied, and the property $\left\|u^{n+1}\right\|^2 \le \|u^n\|^2$ was called monotonicity stability in [43, 42]. The discussions in [43, 35, 42] were focused on explicit RK methods. In the present paper, our framework, including the discrete energy laws and stability results, applies to both general implicit and explicit RK methods.

3.3. On the invertibility of Q. In the above stability analysis, we have assumed that the operator $Q := \mathcal{Q}(\tau L)$ is invertible, i.e., the given (implicit) RK scheme (3.1) is uniquely solvable. We now discuss the invertibility of Q below. Let $\mathbb{A} := \{z_i\}_{i=1}^{s_q}$ denote the set of all the roots of the polynomial $\mathcal{Q}(z)$ in $\mathbb{C}$.

LEMMA 3.11. *Let V be a Hilbert space of either finite or infinite dimensions. We denote by $\sigma(L)$ the spectrum of the operator L. For a fixed $\tau > 0$, the operator $Q := \mathcal{Q}(\tau L)$ is invertible if*

$$\mathbb{A} \cap (\tau\sigma(L)) = \emptyset. \tag{3.20}$$

Furthermore, Q is invertible for any $0 < \tau\|L\| < z_\star$, where $z_\star := \min_{1\le i\le s_q} |z_i| > 0$.

Proof. Observe that $Q = \mathcal{Q}(\tau L) = c(\tau L - z_1 I)(\tau L - z_2 I)\cdots(\tau L - z_{s_p} I)$ for some constant c. Under the condition (3.20), the operators $\tau L - z_i I$, $1 \le i \le s_p$, are all invertible, and thus Q is invertible. Note that $z_\star > 0$ because $\mathcal{Q}(0) = 1 \neq 0$, and $\|L\|$ is larger than or equal to the spectral radius of L. If $\tau\|L\| < z_\star$, then (3.20) holds and Q is invertible. ∎

An improved estimate on τ can be obtained if the space V is finite dimensional (in this case the operator L can be regarded as a matrix). See Theorem 3.12. Similar discussions may also be extended to infinite dimensional spaces if the operator L is compact.

THEOREM 3.12. *Assume that V is finite dimensional and denote $Q := \mathcal{Q}(\tau L)$.*

1. *If all the roots of the polynomial $\mathcal{Q}(z)$ have positive real parts, then Q is invertible for any $\tau > 0$.*
2. *If $\mathcal{Q}(z)$ has at least one root with nonpositive real parts, then Q is invertible for all $0 < \tau < \tau_\star$ with $\tau_\star = \min_{0\neq\sigma\in\sigma(L)} \min_{z_i\in\mathbb{A}_\sigma} |z_i/\sigma|$. Here $\mathbb{A}_\sigma = \{z_i \in \mathbb{A} : \mathrm{Re}(z_i) \le 0, \mathrm{Arg}(z_i) = \mathrm{Arg}(\sigma)\}$, and we define $\tau_\star = +\infty$ if $\mathbb{A}_\sigma = \emptyset$ for all nonzero $\sigma \in \sigma(L)$.*

Proof. For an arbitrary eigenvalue σ of L, let $v + \mathrm{i}w$ be the corresponding eigenvector. Then it can be easily shown that $\langle Lv, v\rangle + \langle Lw, w\rangle = \mathrm{Re}(\sigma)(\|v\|^2 + \|w\|^2) \le 0$, which yields $\mathrm{Re}(\sigma) \le 0$. This means all the eigenvalues of L have nonnegative real parts. Therefore, if $\mathrm{Re}(z_i) > 0$ for all $0 \le i \le s_q$, then the condition (3.20) holds for any τ, and by Lemma 3.11 the operator $Q = \mathcal{Q}(\tau L)$ is always invertible for any $\tau > 0$. If the polynomial $\mathcal{Q}(z)$ has at least one root with nonpositive real parts, then when $0 < \tau < \tau_\star$, we have $\tau\sigma \neq z_i$ for any i and σ, so that (3.20) holds and Q is invertible. ∎

4. Examples on discrete energy laws. This section gives several specific examples of implicit methods to further illustrate the proposed discrete energy law (3.12) in Theorem 3.4. For all the methods studied in this and the next sections, the roots of the corresponding polynomial $\mathcal{Q}(z)$ all have positive real parts, and therefore, by Theorem 3.12, $Q := \mathcal{Q}(\tau L)$ is always invertible for any $\tau > 0$ in these examples.

4.1. Examples of unconditional strong stability. We first use our framework to derive the energy identity for several A-stable implicit RK schemes. For these schemes, the conditions in Theorem 3.5 are satisfied so that the strong stability holds without any time step constraint.

Example 4.1 (Euler backward method). The stability function of this method is $\mathcal{R}(Z) = (I - Z)^{-1}$. Using Lemma 3.3 gives $\mathbf{B} = \mathrm{diag}\{0, -1\}$ and $\boldsymbol{\Upsilon} = (-1) = -\mathbf{U}^\top\mathbf{D}\mathbf{U}$ with $\mathbf{D} = (1)$ and $\mathbf{U} = (1)$. Since $w^n = Q^{-1}u^n = \mathcal{R}(\tau L)u^n = u^{n+1}$, according to Theorem 3.4 we obtain the energy law as

$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 = -\tau^2 \left\|Lu^{n+1}\right\|^2 - \tau \left[\!\left[u^{n+1}\right]\!\right]^2.$$

Example 4.2 (Crank–Nicolson method and implicit midpoint method). The stability functions of these two methods are both $\mathcal{R}(Z) = \left(I - \frac{Z}{2}\right)^{-1}\left(I + \frac{Z}{2}\right)$. By

Lemma 3.3, we have $\mathbf{B} = \mathrm{diag}\{0,0\}$ and $\boldsymbol{\Upsilon} = (-1) = -\mathbf{U}^\top\mathbf{D}\mathbf{U}$ with $\mathbf{D} = (1)$ and $\mathbf{U} = (1)$. According to Theorem 3.4, we obtain the energy identity

$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 = -\tau\left[\!\left[w^n\right]\!\right]^2.$$

Example 4.3 (Qin and Zhang [27]). The Butcher tableau and stability function of this method are

$$\begin{array}{c|cc} \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}, \qquad \mathcal{R}(Z) = \left(I - \frac{Z}{2} + \frac{Z^2}{16}\right)^{-1}\left(I + \frac{Z}{2} + \frac{Z^2}{16}\right).$$

According to Lemma 3.3, we have $\mathbf{B} = \mathrm{diag}\,\{0,0,0\}$ and

$$\boldsymbol{\Upsilon} = \mathrm{diag}\,\{-1,-1/16\} = -\mathbf{U}^\top\mathbf{D}\mathbf{U}, \quad \mathbf{D} = \mathrm{diag}\left\{1,\frac{1}{4}\right\}, \quad \text{and} \quad \mathbf{U} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thanks to Theorem 3.4, we obtain the corresponding energy identity

$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 = -\tau\left[\!\left[w^n\right]\!\right]^2 - \frac{1}{16}\tau^3\left[\!\left[Lw^n\right]\!\right]^2.$$

Example 4.4 (Kraaijevanger and Spijker [22]). The Butcher tableau and corresponding stability function of this method are

$$\begin{array}{c|cc} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{3}{2} & -\frac{1}{2} & 2 \\ \hline & -\frac{1}{2} & \frac{3}{2} \end{array}, \qquad \mathcal{R}(Z) = \left(I - \frac{5Z}{2} + Z^2\right)^{-1}\left(I - \frac{3Z}{2} + \frac{Z^2}{2}\right).$$

According to Lemma 3.3, we have $\mathbf{B} = \mathrm{diag}\left\{0,-3,-\frac{3}{4}\right\}$ and

$$\boldsymbol{\Upsilon} = \begin{pmatrix} -1 & \frac{1}{2} \\ \frac{1}{2} & -\frac{7}{4} \end{pmatrix} = -\mathbf{U}^\top\mathbf{D}\mathbf{U} \quad \text{with} \quad \mathbf{D} = \mathrm{diag}\left\{1,\frac{3}{2}\right\} \quad \text{and} \quad \mathbf{U} = \begin{pmatrix} 1 & -\frac{1}{2} \\ 0 & 1 \end{pmatrix}.$$

By Theorem 3.4, we obtain the discrete energy law as

$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 = -3\tau^2\left\|Lw^n\right\|^2 - \frac{3}{4}\tau^4\left\|L^2w^n\right\|^2 - \tau\left[\!\left[\left(I - \frac{\tau}{2}L\right)w^n\right]\!\right]^2 - \frac{3}{2}\tau^3\left[\!\left[Lw^n\right]\!\right]^2.$$

4.2. Examples of conditional stability. Next, we derive the energy laws for two implicit methods which are not A-stable. Conditional stability can be obtained by Theorem 3.7.

Example 4.5 (weak stability). This example considers the $(0,3)$ Padé approximation with the stability function $\mathcal{R}(Z) = (I - Z + \frac{Z^2}{2} - \frac{Z^3}{6})^{-1}$. This method is $A(\alpha)$-stable with $\alpha \le 88.23^o$; see [39, Chapter IV.3, p. 46]. If applying it to a generic linear seminegative problem (1.1), the unconditional stability would not hold in general. According to Lemma 3.3, we get

$$\mathbf{B} = \mathrm{diag}\left\{0,0,\frac{1}{12},-\frac{1}{36}\right\}, \qquad \boldsymbol{\Upsilon} = \begin{pmatrix} -1 & \frac{1}{2} & -\frac{1}{6} \\ \frac{1}{2} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{6} & \frac{1}{6} & -\frac{1}{12} \end{pmatrix}.$$

Direct calculation shows that $\boldsymbol{\Upsilon}$ has a positive eigenvalue, implying it is not negative semidefinite. But its second-order principle submatrix is negative semidefinite. Moreover, with $\boldsymbol{\Delta} = \operatorname{diag}\left\{0, 0, \frac{1}{36}\right\}$ the matrix $\tilde{\boldsymbol{\Upsilon}} := \boldsymbol{\Upsilon} - \boldsymbol{\Delta}$ is negative semidefinite and admits the following Cholesky type decomposition:

$$\tilde{\boldsymbol{\Upsilon}} = -\tilde{\mathbf{U}}^\top \tilde{\mathbf{D}} \tilde{\mathbf{U}}, \quad \tilde{\mathbf{D}} = \operatorname{diag}\left\{1, \frac{1}{12}, 0\right\}, \quad \tilde{\mathbf{U}} = \begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{6} \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thanks to Theorem 3.4, we obtain the energy identity

$$\begin{aligned} \left\|u^{n+1}\right\|^2 - \|u^n\|^2 = \frac{\tau^4}{12}\left\|L^2 w^n\right\|^2 - \frac{\tau^6}{36}\left\|L^3 w^n\right\|^2 - \tau\left[\!\!\left[\left(I - \frac{\tau}{2}L + \frac{\tau^2}{6}L^2\right) w^n\right]\!\!\right]^2 \\ - \frac{\tau^3}{12}\left[\!\!\left[L(I - \tau L) w^n\right]\!\!\right]^2 + \frac{\tau^5}{36}\left[\!\!\left[L^2 w^n\right]\!\!\right]^2. \end{aligned}$$

Thus $\zeta = 2$ and $\rho = 2$, and by Theorem 3.7, the $(0, 3)$ Padé approximation is weakly(κ) stable with $\kappa = 4$.

Example 4.6 (strong stability). We consider the $(4, 1)$ Padé approximation whose stability function is $R(Z) = (I - \frac{Z}{5})^{-1}(I + \frac{4Z}{5} + \frac{3Z^2}{10} + \frac{Z^3}{15} + \frac{Z^4}{120})$. According to Lemma 3.3, we obtain

$$\mathbf{B} = \operatorname{diag}\left\{0, 0, 0, -\frac{1}{1800}, \frac{1}{14400}\right\} \quad \text{and} \quad \boldsymbol{\Upsilon} = -\begin{pmatrix} 1 & \frac{3}{10} & \frac{1}{15} & \frac{1}{120} \\ \frac{3}{10} & \frac{13}{75} & \frac{9}{200} & \frac{1}{150} \\ \frac{1}{15} & \frac{9}{200} & \frac{1}{75} & \frac{1}{400} \\ \frac{1}{120} & \frac{1}{150} & \frac{1}{400} & \frac{1}{1800} \end{pmatrix}.$$

Direct calculation shows that $\boldsymbol{\Upsilon}$ has a positive eigenvalue, implying it is not negative semidefinite. But its third-order principle submatrix is negative semidefinite. Moreover, with $\boldsymbol{\Delta} = \operatorname{diag}\left\{0, 0, 0, \frac{1}{14400}\right\}$, the matrix $\tilde{\boldsymbol{\Upsilon}} := \boldsymbol{\Upsilon} - \boldsymbol{\Delta}$ is negative semidefinite and admits the following Cholesky type decomposition:

$$\tilde{\boldsymbol{\Upsilon}} = -\tilde{\mathbf{U}}^\top \tilde{\mathbf{D}} \tilde{\mathbf{U}}, \quad \tilde{\mathbf{D}} = \operatorname{diag}\left\{1, \frac{1}{12}, \frac{1}{720}, 0\right\}, \quad \tilde{\mathbf{U}} = \begin{pmatrix} 1 & \frac{3}{10} & \frac{1}{15} & \frac{1}{120} \\ 0 & 1 & \frac{3}{10} & \frac{1}{20} \\ 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

According to Theorem 3.4, we have the following energy law:

$$\begin{aligned} &\left\|u^{n+1}\right\|^2 - \|u^n\|^2 \\ &= -\frac{\tau^6}{1800}\left\|L^3 w^n\right\|^2 + \frac{\tau^8}{14400}\left\|L^4 w^n\right\|^2 - \tau\left[\!\!\left[\left(I + \frac{3\tau}{10}L + \frac{\tau^2}{15}L^2 + \frac{\tau^3}{120}L^3\right) w^n\right]\!\!\right]^2 \\ &\quad - \frac{\tau^3}{12}\left[\!\!\left[L\left(I + \frac{3\tau}{10}L + \frac{\tau^2}{20}L^2\right) w^n\right]\!\!\right]^2 - \frac{\tau^5}{720}\left[\!\!\left[L^2\left(I + \frac{\tau}{2}L\right) w^n\right]\!\!\right]^2 + \frac{\tau^7}{14400}\left[\!\!\left[L^3 w^n\right]\!\!\right]^2. \end{aligned}$$

This implies $\zeta = 3$, $\rho = 3$, and $\beta_\zeta < 0$. We conclude the conditional strong stability from Theorem 3.7.

5. Unified energy law for general diagonal Padé approximations. In this section, we apply the proposed framework to derive the unified discrete energy law for general diagonal Padé approximations of arbitrary order. *The establishment of such an energy law will be based on a highly technical Cholesky type decomposition of a family of complicated matrices, whose discovery and proof are extremely nontrivial.*

It was shown in [1, Lemma 7] for any diagonal Padé approximations that all the zeros of the polynomial $\mathcal{Q}(z)$ have positive real parts, so that $Q = \mathcal{Q}(\tau L)$ is always invertible for any $\tau > 0$ by Theorem 3.12. For the (s, s) diagonal Padé approximation, the stability function is given by (3.2) and (3.3) with the coefficients in (3.3) defined as

$$\theta_i = (-1)^i \vartheta_i = \frac{s!}{(2s)!} \frac{(2s-i)!}{i!(s-i)!}. \tag{5.1}$$

Thus we have $\alpha_{i,j} = \theta_i\theta_j - \vartheta_i\vartheta_j = (1-(-1)^{i+j})\theta_i\theta_j$. According to Lemma 3.3, the matrix $\mathbf{B} = \mathrm{diag}(\{\beta_k\}_{k=0}^{s}) = \mathbf{O}$ because

$$\beta_k = \sum_{\ell=\max\{0,2k-s\}}^{\min\{2k,s\}} \alpha_{\ell,2k-\ell}(-1)^{k-\ell} = \sum_{\ell=\max\{0,2k-s\}}^{\min\{2k,s\}} (1-(-1)^{2k})\theta_\ell\theta_{2k-\ell}(-1)^{k-\ell} = 0,$$

and the symmetric matrix $\boldsymbol{\Upsilon} = (\gamma_{i,j})_{i,j=0}^{s-1}$ is computed by

$$\gamma_{i,j} = \sum_{\ell=\max\{0,i+j+1-s\}}^{\min\{i,j\}} (-1)^{\min\{i,j\}+1-\ell}\Big(1-(-1)^{i+j+1}\Big)\theta_\ell\theta_{i+j+1-\ell}$$

(5.2)

$$= \Big((-1)^i + (-1)^j\Big) \sum_{\ell=\max\{0,i+j+1-s\}}^{\min\{i,j\}} (-1)^{\ell+1}\theta_\ell\theta_{i+j+1-\ell}$$

(5.3)

$$= \Big((-1)^i+(-1)^j\Big)\left(\frac{s!}{(2s)!}\right)^2 \sum_{\ell=\max\{0,i+j+1-s\}}^{\min\{i,j\}} (-1)^{\ell+1} \frac{(2s-\ell)!}{\ell!(s-\ell)!} \frac{(2s-i-j-1+\ell)!}{(i+j+1-\ell)!(s-i-j-1+\ell)!}.$$

In order to establish the energy identity, the key step is to judge the negative semi-definiteness of the above matrix $\boldsymbol{\Upsilon}$ and construct its Cholesky type decomposition. For an arbitrary $s \in \mathbb{Z}^+$, this is indeed a highly challenging task, because the structures of $\boldsymbol{\Upsilon}$ are extremely complicated and all its elements (5.3) involve complex summations of several factorial products.

After careful investigation, we find the unified explicit form of the Cholesky type decomposition of $\boldsymbol{\Upsilon}$, as stated in Theorem 5.1.

THEOREM 5.1 (constructive matrix decomposition). *For any $s \in \mathbb{Z}^+$, the symmetric matrix $\boldsymbol{\Upsilon}$ defined by* (5.3) *is always negative definite. Furthermore, it has the Cholesky type decomposition in the following unified explicit form:*

$$\boldsymbol{\Upsilon} = -\mathbf{U}^\top \widehat{\mathbf{D}} \mathbf{U}, \tag{5.4}$$

where $\widehat{\mathbf{D}} = \mathrm{diag}(\{\widehat{d}_k\}_{k=0}^{s-1})$ with $\widehat{d}_k = \frac{(k!)^2}{(2k)!(2k+1)!}$, and $\mathbf{U} = (\mu_{i,j})_{i,j=0}^{s-1}$ is an upper triangular matrix with

(5.5)

$$\mu_{i,j} := \begin{cases} \dfrac{s!}{(2s)!}\dfrac{(2i+1)!}{i!(i+j+1)!}\dfrac{(2s+i-j)!}{(s-1-j)!}\dfrac{(s-1-\frac{i+j}{2})!(\frac{i+j}{2})!}{(s-\frac{j-i}{2})!(\frac{j-i}{2})!} & \text{if } i \le j \text{ and } i \equiv j \pmod 2, \\ 0 & \text{otherwise.} \end{cases}$$

The proof of Theorem 5.1 is very technical and will be given in subsection 5.3 for better readability.

5.1. Unified discrete energy law and unconditional stability. Combining Theorem 5.1 with Theorem 3.4, we immediately obtain the discrete energy laws of all the diagonal Padé approximations in a unified form.

THEOREM 5.2 (unified energy law and unconditional stability). *For any $s \in \mathbb{Z}^+$, the (s,s) diagonal Padé approximation for general linear seminegative system* (1.1) *admits the discrete energy law*

$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 = -\sum_{k=0}^{s-1} \widehat{d}_k \tau^{2k+1} \left[\!\left[L^k u^{(k)} \right]\!\right]^2, \tag{5.6}$$

where $\widehat{d}_k = \frac{(k!)^2}{(2k)!(2k+1)!}$ *and*

$$u^{(k)} := \sum_{j=k}^{s-1} \mu_{k,j} (\tau L)^{j-k} Q^{-1} u^n \tag{5.7}$$

with $\mu_{k,j}$ defined by (5.5). *The energy law* (5.6) *implies $\left\|u^{n+1}\right\| \le \|u^n\|$ for all $\tau > 0$, which means all diagonal Padé approximations are unconditionally strongly stable for general linear seminegative systems.*

Remark 5.3 (comparison with the algebraic stability analysis). Note that the diagonal Padé approximations correspond to the stability functions of the Gauss methods, the Lobatto IIIA/IIIB methods, etc.; see [39, Table 5.13, p. 82]. Therefore, the unconditional strong stability in Theorem 5.2 is consistent with the classical results [15, 18] on algebraic stability of the Gauss methods in the special case of (1.1). In fact, one can derive another (different) energy law via the algebraic stability analysis of the Gauss methods. Suppose the Gauss methods for (1.1) can be written as

$$u^{n+1} = u^n + \tau \sum_{k=1}^{s} b_k L u_G^{(k)}, \quad \text{and} \quad u_G^{(k)} = u^n + \tau \sum_{j=1}^{s} a_{k,j} L u_G^{(j)} \quad \forall 1 \le k \le s, \tag{5.8}$$

where $b_k \ge 0$ for all $1 \le k \le s$ and $\mathbf{M} = (m_{kj}) = (b_k a_{k,j} + b_j a_{j,k} - b_k b_j)_{k,j=1}^s = \mathbf{O}$ for the Gauss methods [18, section 5.2, pp. 82–83]. Then applying the energy argument in [18, Proof of Theorem 5.2] or [39, equation (12.7)] to (5.8), we can obtain a different energy identity

$$\left\|u^{n+1}\right\|^2 - \|u^n\|^2 = -\sum_{k=1}^{s} b_k \tau \left[\!\left[u_G^{(k)} \right]\!\right]^2. \tag{5.9}$$

As we can see, our energy identity (5.6) and the above identity (5.9), which are derived in very different ways, both imply a consistent fact—the unconditional strong stability of diagonal Padé approximations. However, these two energy identities (5.6) and (5.9) have quite different structures. Specifically, in (5.9), the energy dissipation is represented as $\mathcal{O}(\tau)$ terms associated with the stage variables of the Gauss methods, whereas in (5.6), our energy law resembles the continuous case in Theorem 2.1 and the energy dissipation is sorted out as $\mathcal{O}(\tau)$, $\mathcal{O}(\tau^3)$, $\ldots$, $\mathcal{O}(\tau^{2s-1})$ terms. This distinct feature of (5.6) helps us to establish a close connection between the continuous and discrete energy laws and provides some new insights on the intrinsic mechanisms. See subsection 5.2 for further discussions.

5.2. Connections between continuous and discrete energy laws. Having found the above unified discrete energy law (5.6), we are now in position to explore the connections between the continuous energy law (2.1) in Theorem 2.1 and the discrete energy law (5.6) in Theorem 5.2.

In fact, the discrete energy law (5.6) of the (s,s) diagonal Padé approximation is a truncated approximation to the continuous energy law (2.1). It is clearly seen that the continuous and discrete laws share the same expansion coefficients $\widehat{d}_k$ of the first s terms. Although the quantity $u^{(k)}$ in (5.7) is not exactly equal to $\widehat{u}^{(k)}$ in (2.2), they actually match up to high order. Notice that the series $u^{(k)}$ in (5.7) is expanded in terms of $w^n = Q^{-1}u^n$, while $\widehat{u}^{(k)}$ in (2.2) is expanded in terms of $u(t^n)$. For ease of comparison, we can either reformulate $\widehat{u}^{(k)}$ in a similar form as $u^{(k)}$ (see Theorem 5.5) or rewrite $u^{(k)}$ in a similar form as $\widehat{u}^{(k)}$ (see Theorem 5.6). In order to rigorously show these two theorems, we need the important combinatorial identity in Lemma 5.4, whose proof is provided in Appendix D.

LEMMA 5.4. *For any $i,j \in \mathbb{N}$ and $s \in \mathbb{Z}^+$ with $0 \le i \le j \le s-1$, it holds that*

$$\sum_{\ell=0}^{j-i} \binom{s-\ell}{j-\ell}^{-1} \binom{2s-\ell}{j-i-\ell} \binom{i+j+1}{\ell} (-1)^\ell = \begin{cases} \dfrac{(s-1-\frac{i+j}{2})!(\frac{i+j}{2})!}{(s-\frac{j-i}{2})!(\frac{j-i}{2})!}(s-j) & \text{if } i \le j \text{ and } i \equiv j \pmod 2, \\ 0 & \text{otherwise.} \end{cases}$$

THEOREM 5.5. *Suppose $u^n = u(t^n)$. The series $\widehat{u}^{(k)}$ in (2.2) can be equivalently rewritten as*

$$\widehat{u}^{(k)} = \sum_{j=k}^{\infty} \bar{\mu}_{k,j} (\tau L)^{j-k} Q^{-1} u^n, \tag{5.10}$$

where $\bar{\mu}_{k,j} := \sum_{\ell=\max\{j-s,k\}}^{j} \widehat{\mu}_{k,\ell}\vartheta_{j-\ell}$. Moreover, the coefficients $\bar{\mu}_{k,j}$ exactly coincide with those in (5.7), namely, $\bar{\mu}_{k,j} = \mu_{k,j}$ for $k \le j \le s-1$.

Proof. Substituting $u^n = Qw^n = \sum_{k=0}^{s} \vartheta_k (\tau L)^k w^n$ into (2.2), we obtain

$$\begin{aligned} \widehat{u}^{(i)} &= \sum_{\ell=i}^{\infty} \widehat{\mu}_{i,\ell} (\tau L)^\ell \left(\sum_{k=0}^{s} \vartheta_k (\tau L)^k \right) w^n = \sum_{\ell=i}^{\infty} \sum_{k=0}^{s} \widehat{\mu}_{i,\ell}\vartheta_k (\tau L)^{\ell+k} w^n \\ &= \sum_{\ell=i}^{\infty} \sum_{j=\ell}^{\ell+s} \widehat{\mu}_{i,\ell}\vartheta_{j-\ell} (\tau L)^j w^n = \sum_{j=i}^{\infty} \left(\sum_{\ell=\max\{j-s,i\}}^{j} \widehat{\mu}_{i,\ell}\vartheta_{j-\ell} \right) (\tau L)^j w^n =: \sum_{j=i}^{\infty} \bar{\mu}_{i,j} (\tau L)^j w^n. \end{aligned}$$

Recall the definitions of $\widehat{\mu}_{i,j}$ and ϑ_i in (2.3) and (5.1), respectively. Substituting them into $\bar{\mu}_{i,j}$, we have

$$\begin{aligned} \bar{\mu}_{i,j} &= \sum_{\ell=\max\{j-s,i\}}^{j} \frac{(2i+1)!\ell!}{i!(\ell-i)!(\ell+i+1)!} \frac{s!}{(2s)!} \frac{(2s-(j-\ell))!}{(j-\ell)!(s-(j-\ell))!} (-1)^{j-\ell} \\ &= \frac{s!}{(2s)!} \frac{(2i+1)!}{i!} \frac{(2s+i-j)!}{(s-j)!} \sum_{\ell=\max\{j-s,i\}}^{j} \frac{(-1)^{j-\ell}\ell!}{(\ell-i)!(\ell+i+1)!} \frac{(s-j)!}{(2s+i-j)!} \frac{(2s-(j-\ell))!}{(j-\ell)!(s-(j-\ell))!} \\ &= \frac{s!}{(2s)!} \frac{(2i+1)!}{i!(i+j+1)!} \frac{(2s+i-j)!}{(s-j)!} \sum_{\ell=\max\{j-s,i\}}^{j} (-1)^{j-\ell} \binom{s+\ell-j}{\ell}^{-1} \binom{2s+\ell-j}{\ell-i} \binom{i+j+1}{j-\ell} \\ &= \frac{s!}{(2s)!} \frac{(2i+1)!}{i!(i+j+1)!} \frac{(2s+i-j)!}{(s-j)!} \sum_{\ell=0}^{\min\{j-i,s\}} \binom{s-\ell}{j-\ell}^{-1} \binom{2s-\ell}{j-i-\ell} \binom{i+j+1}{\ell} (-1)^\ell. \end{aligned}$$

Note that when $i \le j \le s-1$, we have $\min\{j-i, s\} = j-i$. Using the combinatorial identity in Lemma 5.4, we obtain $\bar{\mu}_{i,j} = \mu_{i,j}$ for $0 \le i, j \le s-1$. The proof is completed. ∎

THEOREM 5.6. *The series $u^{(k)}$ in (5.7) can be equivalently reformulated as*

$$u^{(k)} = \sum_{j=k}^{s-1} \widehat{\mu}_{k,j} (\tau L)^{j-k} I_j u^n,$$

where $I_j := Q_j Q^{-1}$ with $Q_j := \sum_{i=0}^{s-1-j} \vartheta_i (\tau L)^i$ denoting the $(s-1-j)$th order truncation of Q.

Proof. According to Theorem 5.5, we have $\bar{\mu}_{i,j} = \mu_{i,j}$ for $0 \le i, j \le s-1$. In this case, $\max\{j-s, i\} = i$ and thus $\mu_{i,j} = \bar{\mu}_{i,j} = \sum_{\ell=\max\{j-s,i\}}^{j} \widehat{\mu}_{i,\ell} \vartheta_{j-\ell} = \sum_{\ell=i}^{j} \widehat{\mu}_{i,\ell} \vartheta_{j-\ell}$. Substituting this into (5.7) gives

$$\begin{aligned} u^{(i)} &= \sum_{j=i}^{s-1} \left(\sum_{\ell=i}^{j} \widehat{\mu}_{i,\ell} \vartheta_{j-\ell} \right) (\tau L)^{j-i} Q^{-1} u^n = \sum_{\ell=i}^{s-1} \sum_{j=\ell}^{s-1} \widehat{\mu}_{i,\ell} \vartheta_{j-\ell} (\tau L)^{j-i} Q^{-1} u^n \\ &= \sum_{\ell=i}^{s-1} \sum_{j=0}^{s-1-\ell} \widehat{\mu}_{i,\ell} \vartheta_j (\tau L)^{j+\ell-i} Q^{-1} u^n = \sum_{\ell=i}^{s-1} \widehat{\mu}_{i,\ell} (\tau L)^{\ell-i} \left(\sum_{j=0}^{s-1-\ell} \vartheta_j (\tau L)^j \right) Q^{-1} u^n, \end{aligned}$$

which completes the proof. ∎

Remark 5.7. As a direct corollary of the above energy laws, Theorem 5.5 together with Theorems 2.1 and 5.2 gives $\left(\left\| u(t^{n+1}) \right\|^2 - \| u(t^n) \|^2 \right) - \left(\left\| u^{n+1} \right\|^2 - \| u^n \|^2 \right) = \mathcal{O}(\tau^{2s+1})$, which implies for a fixed $T = n\tau$ that the total energy dissipation accuracy $\Delta E := (\| u(t^n) \|^2 - \left\| u(t^0) \right\|^2) - (\| u^n \|^2 - \left\| u^0 \right\|^2) = \mathcal{O}(\tau^{2s})$. This is consistent with the accuracy of the diagonal Padé approximations (the Gauss methods), as expected.

Remark 5.8. Combining Theorem 5.2 with Theorem 5.6, we can derive the following precise characterization on the operator $\mathcal{R}(\tau L)$:

(5.11)
$$(\mathcal{R}(\tau L))^\top \mathcal{R}(\tau L) - I = \sum_{k=0}^{s-1} \widehat{d}_k \tau^{2k+1} U_k^\top (L^\top + L) U_k \le O, \quad \text{with } U_k = L^k \sum_{j=k}^{s-1} \widehat{\mu}_{k,j} (\tau L)^{j-k} I_j,$$

where $\widehat{d}_k$ and $\widehat{\mu}_{k,j}$ are defined in (2.3), and I_j is defined in Theorem 5.6. Note that the operator $\mathcal{R}(\tau L)$ is the discrete approximation to the operator $e^{\tau L}$. The identity (5.11) on $\mathcal{R}(\tau L)$ is exactly the discrete counterpart of the identity (2.5) on $e^{\tau L}$ of the continuous case.

In summary, our above analyses clearly demonstrate the unity of continuous and discrete objects.

5.3. Proof of Theorem 5.1. The discovery and proof of Theorem 5.1 are highly nontrivial and challenging. Our proof is very technical and relies on several lemmas and constructive identities.

Note that the negative definiteness of $\boldsymbol{\Upsilon}$ is implied by the existence of the Cholesky type decomposition (5.4) with positive $\widehat{d}_k$ for all k. Therefore, we only need to prove the identity (5.4) for any $s \in \mathbb{Z}^+$. Define $\mathbf{F}(s) := \boldsymbol{\Upsilon} + \mathbf{U}^\top \widehat{\mathbf{D}} \mathbf{U}$. Then the goal is to show that the matrix-valued function $\mathbf{F}(s) \equiv \mathbf{O}$ is identically zero for all $s \in \mathbb{Z}^+$.

Let $\mathcal{F}_{p,q}(s)$ denote the (p,q) element of $\mathbf{F}(s)$. In order to clearly show the dependence of $\mathcal{F}_{p,q}(s)$ on s, we will equivalently reformulate it with some new notation. First, we introduce

$$\theta_0^{(s)} := 1, \qquad \theta_i^{(s)} := \frac{1}{i!}\,\frac{s(s-1)\cdots(s-i+1)}{2s(2s-1)\cdots(2s-i+1)}, \quad i \in \mathbb{Z}^+, \tag{5.12}$$

which satisfy $\theta_i^{(s)} = \theta_i$ for $0 \le i \le s$ and $\theta_i^{(s)} = 0$ for $s < i < 2s$. Furthermore, we define

$$\gamma_{p,q}^{(s)} := [(-1)^p + (-1)^q] \sum_{i=0}^{\min\{p,q\}} (-1)^{i+1}\theta_i^{(s)}\theta_{p+q+1-i}^{(s)}, \qquad p,q \in \mathbb{N}. \tag{5.13}$$

Note that $\theta_{p+q+1-i}^{(s)} = 0$ for $p+q+1-i > s$, which along with (5.2) implies

$$\gamma_{p,q}^{(s)} = \gamma_{p,q}, \quad 0 \le p,q \le s-1. \tag{5.14}$$

For $i,j \in \mathbb{Z}^+$, we define

$$\nu_{i,j}^{(s)} := \begin{cases} \dfrac{s!}{(2s)!}\,\dfrac{2\sqrt{2i-1}}{(i+j)!}\,\dfrac{(2s+i-j)!\,\left(s-\frac{i+j}{2}\right)!\,\left(\frac{i+j}{2}\right)!}{(s-j)!\,\left(s-\frac{j-i}{2}\right)!\,\left(\frac{j-i}{2}\right)!} & \text{if } i \le j \text{ and } i \equiv j \pmod 2, \\ 0 & \text{otherwise.} \end{cases} \tag{5.15}$$

One can verify that $\nu_{i,j}^{(s)} = \sqrt{d_{i-1}}\mu_{i-1,j-1}$ for $1 \le i,j \le s$. Therefore, for $1 \le p,q \le s$, $\mathcal{F}_{p,q}(s)$ can be equivalently reformulated as

$$\mathcal{F}_{p,q}(s) = \gamma_{p-1,q-1}^{(s)} + \sum_{i=1}^{\min\{p,q\}} \nu_{i,p}^{(s)}\nu_{i,q}^{(s)} \qquad \forall s \in \mathbb{Z} \tag{5.16}$$

We have the following two crucial observations.

Observation 5.9. For any fixed $p,q \in \mathbb{Z}^+$, the function $\mathcal{F}_{p,q}(s)$ in (5.16) is a rational function of s.

Proof. For any fixed $i \in \mathbb{N}$, the function $\theta_i^{(s)}$ defined in (5.12) is a rational function of s, and thus for any fixed $p,q \in \mathbb{N}$, the function $\gamma_{p,q}^{(s)}$ is also a rational function of s. Note that for any fixed $i,j \in \mathbb{Z}^+$, $\nu_{i,j}^{(s)}$ in (5.15) can be easily rewritten as a rational function of s. Therefore, for any fixed $p,q \in \mathbb{Z}^+$, all the terms in (5.16) are rational functions of s, and thus $\mathcal{F}_{p,q}(s)$ is also a rational function of s. ∎

Observation 5.10. All elements of $\mathbf{F}(s)$ are rational functions of s. Recall that a rational function vanishes at only finite points unless it is identically zero. Therefore, if we can prove that all elements $\mathcal{F}_{p,q}(s)$ vanish for all s on an uncountable set $\widehat{\mathbb{R}}$, then it forces $\mathbf{F}(s) \equiv \mathbf{O}$ for all $s \in \mathbb{Z}^+$.

For convenience, hereafter the factorial is extended to represent the gamma function $\Gamma(x+1)$, namely,

$$x! := \Gamma(x+1) \qquad \forall x \in \mathbb{R} \setminus \mathbb{Z}^-.$$

In our following lemmas and proofs, we will introduce some intermediate quantities that are also rational functions of s, whose denominators may vanish at $\{0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \dots\}$. To avoid potential singularity of dividing a zero denominator, we will extend the domain of s from $\mathbb{Z}^+$ to $\mathbb{R}$ but excluding all potential singular points. More specifically, we will prove the following proposition.

PROPOSITION 5.11. *For all $p, q \in \mathbb{Z}^+$, the rational function $F_{p,q}(s)$ vanishes for all $s \in \widehat{\mathbb{R}}$, namely,*

$$\gamma^{(s)}_{p-1,q-1} + \sum_{i=1}^{\min\{p,q\}} \nu^{(s)}_{i,p}\nu^{(s)}_{i,q} = 0 \qquad \forall p, q \in \mathbb{Z}^+, \qquad \forall s \in \widehat{\mathbb{R}}, \tag{5.17}$$

where

$$\widehat{\mathbb{R}} := \{x \in \mathbb{R} : \ 2x \notin \mathbb{Z}\} = \mathbb{R} \setminus \left\{0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \dots\right\}. \tag{5.18}$$

The proof of Proposition 5.11 relies on several lemmas in subsection 5.4 and will be given in subsection 5.5. Note that the set $\widehat{\mathbb{R}}$ defined in (5.18) is uncountable. Based on Observations 5.9 and 5.10 and the above arguments, once we prove Proposition 5.11, then we immediately obtain (5.4) for all $s \in \mathbb{Z}^+$ and complete the proof of Theorem 5.1.

5.4. Lemmas. This section gives several important lemmas, which pave the way to proving Proposition 5.11. First, we introduce the rising factorial (sometimes also called the Pochhammer symbol in the theory of hypergeometric functions), defined by

$$(x)_0 := 1, \qquad (x)_n := x(x+1)\cdots(x+n-1) = \prod_{k=0}^{n-1}(x+k), \quad n \in \mathbb{Z}^+, \tag{5.19}$$

for any $x \in \mathbb{R}$. Note that

$$(x)_n \neq 0 \qquad \forall x \notin \mathbb{Z} \quad \forall n \in \mathbb{N}.$$

Lemma 5.12 gives three useful identities related to the Pochhammer symbol, whose proofs are presented in Appendix E.

LEMMA 5.12. *The following identities hold:*

$$(x+n)! = x!(x+1)_n \qquad \forall x \in \mathbb{R}, \quad \forall n \in \mathbb{N}, \tag{5.20}$$

$$(x)_n = 2^n \left(\frac{x}{2}\right)_{\lceil \frac{n}{2} \rceil} \left(\frac{x+1}{2}\right)_{\lfloor \frac{n}{2} \rfloor} \qquad \forall x \in \mathbb{R}, \quad \forall n \in \mathbb{N}, \tag{5.21}$$

$$\frac{(x+i)!}{(x-j)!} = (-1)^j(-x)_j(x+1)_i \qquad \forall x \in \mathbb{R} \setminus \{j-1, j-2, \dots\}, \quad \forall i, j \in \mathbb{N}. \tag{5.22}$$

Note for any fixed $i, j \in \mathbb{Z}^+$ that $\nu^{(s)}_{i,j}$ is also a rational function of s. We now establish the relations between $\nu^{(s)}_{i,j}$ and $\theta^{(s)}_j$.

LEMMA 5.13. *For any $i, j \in \mathbb{Z}^+$ and any $s \in \widehat{\mathbb{R}}$, we have*

$$\nu^{(s)}_{2i,2j} = 2\sqrt{4i-1}\frac{\left(s+\frac{1}{2}-j\right)_i (-j)_i}{(j-s)_i \left(\frac{1}{2}+j\right)_i}\theta^{(s)}_{2j}, \tag{5.23}$$

$$\nu^{(s)}_{2i-1,2j-1} = 2\sqrt{4i-3}\frac{\left(s+\frac{3}{2}-j\right)_{i-1} (1-j)_{i-1}}{(j-s)_{i-1} \left(\frac{1}{2}+j\right)_{i-1}}\theta^{(s)}_{2j-1}. \tag{5.24}$$

The proof of Lemma 5.13 is put in Appendix F.

For $p, q \in \mathbb{Z}^+$, define the following two sequences of rational functions of s: for $n = 0, 1, \ldots,$

$$\varphi_n(s;p,q) := \frac{\left(s+\frac{3}{2}-p\right)_n (1-p)_n \left(s+\frac{1}{2}-q\right)_n (-q)_n}{(p-s+1)_n \left(p+\frac{3}{2}\right)_n (q-s+1)_n \left(q+\frac{3}{2}\right)_n}, \tag{5.25}$$

$$\phi_n(s;p,q) := \varphi_n(s;p,q)\frac{\mathcal{C}^{1,s}_{n,p,q} + \mathcal{C}^{2,s}_{n,p,q}}{(s-p)(1+2p)(s-q)(1+2q)} \tag{5.26}$$

with

$$\begin{aligned}\mathcal{C}^{1,s}_{n,p,q} &:= (4n+3)(1+s-2p)(q-n)(1+2s+2n-2q),\\ \mathcal{C}^{2,s}_{n,p,q} &:= (4n+1)(s-2q)(1+2p+2n)(s-p-n).\end{aligned}$$

Notice that for all $n \geq p$, we have $(1-p)_n = 0$, so that

$$\varphi_n(s;p,q) = 0, \quad \phi_n(s;p,q) = 0 \qquad \forall n \geq p. \tag{5.27}$$

LEMMA 5.14. *For any $s \in \widehat{\mathbb{R}}$, it holds that*

$$\nu^{(s)}_{2i-1,2p-1}\nu^{(s)}_{2i-1,2q+1} + \nu^{(s)}_{2i,2p}\nu^{(s)}_{2i,2q} = 2\theta^{(s)}_{2p-1}\theta^{(s)}_{2q}\phi_{i-1}(s;p,q) \qquad \forall i,p,q \in \mathbb{Z}^+. \tag{5.28}$$

Proof. Denote $n = i-1$. Using Lemma 5.13 gives

$$\begin{aligned}\nu^{(s)}_{2i-1,2p-1}\nu^{(s)}_{2i-1,2q+1} &\overset{(5.24)}{=} 4(4i-3)\theta^{(s)}_{2p-1}\theta^{(s)}_{2q+1}\frac{\left(s+\frac{3}{2}-p\right)_{i-1}(1-p)_{i-1}}{(p-s)_{i-1}\left(p+\frac{1}{2}\right)_{i-1}}\frac{\left(s+\frac{1}{2}-q\right)_{i-1}(-q)_{i-1}}{(q+1-s)_{i-1}\left(q+\frac{3}{2}\right)_{i-1}}\\ &= 4(4n+1)\theta^{(s)}_{2p-1}\theta^{(s)}_{2q+1}\frac{(p+1-s)_n\left(p+\frac{3}{2}\right)_n}{(p-s)_n\left(p+\frac{1}{2}\right)_n}\varphi_n(s;p,q)\\ &= 4(4n+1)\theta^{(s)}_{2p-1}\frac{\theta^{(s)}_{2q}(s-2q)}{(2s-2q)(2q+1)}\frac{(p-s+n)\left(p+n+\frac{1}{2}\right)}{(p-s)\left(p+\frac{1}{2}\right)}\varphi_n(s;p,q)\\ &= 2\theta^{(s)}_{2p-1}\theta^{(s)}_{2q}\frac{(4n+1)(s-2q)(1+2p+2n)(s-p-n)}{(s-p)(1+2p)(s-q)(1+2q)}\varphi_n(s;p,q).\end{aligned}$$

Applying Lemma 5.13 and using $(x)_{n+1} = (x)_n(x+n)$ and $(x)_{n+1} = (x+1)_n x$, we can deduce

$$\begin{aligned}\nu^{(s)}_{2i,2p}\nu^{(s)}_{2i,2q} &\overset{(5.23)}{=} 4(4i-1)\theta^{(s)}_{2p}\theta^{(s)}_{2q}\frac{\left(s+\frac{1}{2}-p\right)_i(-p)_i}{(p-s)_i\left(p+\frac{1}{2}\right)_i}\frac{\left(s+\frac{1}{2}-q\right)_i(-q)_i}{(q-s)_i\left(q+\frac{1}{2}\right)_i}\\ &= 4(4n+3)\theta^{(s)}_{2p-1}\frac{\theta^{(s)}_{2q}(s-2p+1)}{2p(2s-2p+1)}\frac{\left(s+\frac{1}{2}-p\right)_{n+1}(-p)_{n+1}}{(p-s)_{n+1}\left(p+\frac{1}{2}\right)_{n+1}}\frac{\left(s+\frac{1}{2}-q\right)_{n+1}(-q)_{n+1}}{(q-s)_{n+1}\left(q+\frac{1}{2}\right)_{n+1}}\\ &= 2\theta^{(s)}_{2p-1}\theta^{(s)}_{2q}\frac{(4n+3)(1+s-2p)(q-n)(1+2s+2n-2q)}{(s-p)(1+2p)(s-q)(1+2q)}\varphi_n(s;p,q).\end{aligned}$$

Combining the above two equations gives (5.28) and completes the proof. ∎

LEMMA 5.15. *For $p, q \in \mathbb{Z}^+$, define a sequence of rational functions of s: for $n = 0, 1, \ldots,$*

$$\Phi_n(s;p,q) := \frac{\mathcal{C}^{3,s}_{n,p,q}}{(s-p)(1+2p)(s-q)(1+2q)}\varphi_n(s;p,q) \tag{5.29}$$

with $\mathcal{C}^{3,s}_{n,p,q} := (n+p-s)(1+2p+2n)(n+q-s)(1+2q+2n)$. *Then, for any* $s \in \widehat{\mathbb{R}}$ *and* $p, q \in \mathbb{Z}^+$, *we have*

$$\Phi_0(s;p,q) = 1, \tag{5.30}$$

$$\Phi_n(s;p,q) = 0 \qquad \forall n \geq p, \tag{5.31}$$

$$\Phi_{n+1}(s;p,q) - \Phi_n(s;p,q) = -\phi_n(s;p,q) \qquad \forall n \in \mathbb{N}. \tag{5.32}$$

Proof. Proof of (5.30). Because $(x)_0 = 1$, we have $\varphi_0(s;p,q) = 1$. Then by $\mathcal{C}^{3,s}_{0,p,q} = (p-s)(1+2p)(q-s)(1+2q)$, we obtain $\Phi_0(s;p,q) = \varphi_0(s;p,q) = 1$.

Proof of (5.31). Recall (5.27) shows $\varphi_n(s;p,q) = 0$ for all $n \geq p$. This immediately leads to (5.31).

Proof of (5.32). Utilizing the relation $(x)_{n+1} = (x)_n(x+n)$ gives

$$\begin{aligned}\varphi_{n+1}(s;p,q) &= \frac{\left(s+\frac{3}{2}-p+n\right)(1-p+n)\left(s+\frac{1}{2}-q+n\right)(n-q)}{(p-s+1+n)\left(p+\frac{3}{2}+n\right)(q-s+1+n)\left(q+\frac{3}{2}+n\right)}\varphi_n(s;p,q)\\ &=: \mathcal{C}^{4,s}_{n,p,q}\varphi_n(s;p,q).\end{aligned}$$

It follows that

$$\Phi_{n+1}(s;p,q) = \frac{\mathcal{C}^{3,s}_{n+1,p,q}\varphi_{n+1}(s;p,q)}{(s-p)(1+2p)(s-q)(1+2q)} = \frac{\mathcal{C}^{3,s}_{n+1,p,q}\mathcal{C}^{4,s}_{n,p,q}\varphi_n(s;p,q)}{(s-p)(1+2p)(s-q)(1+2q)}$$

with $\mathcal{C}^{3,s}_{n+1,p,q}\mathcal{C}^{4,s}_{n,p,q} = (2n+2s-2p+3)(n-p+1)(2n+2s-2q+1)(n-q)$. By direct calculations, we observe that the identity $\mathcal{C}^{3,s}_{n+1,p,q}\mathcal{C}^{4,s}_{n,p,q} - \mathcal{C}^{3,s}_{n,p,q} = -\mathcal{C}^{1,s}_{n,p,q} - \mathcal{C}^{2,s}_{n,p,q}$ always holds, which leads to

$$\begin{aligned}\Phi_{n+1}(s;p,q) - \Phi_n(s;p,q) &= \frac{\mathcal{C}^{3,s}_{n+1,p,q}\mathcal{C}^{4,s}_{n,p,q} - \mathcal{C}^{3,s}_{n,p,q}}{(s-p)(1+2p)(s-q)(1+2q)}\varphi_n(s;p,q)\\ &= \frac{-\mathcal{C}^{1,s}_{n,p,q} - \mathcal{C}^{2,s}_{n,p,q}}{(s-p)(1+2p)(s-q)(1+2q)}\varphi_n(s;p,q) = -\phi_n(s;p,q). \quad \square\end{aligned}$$

LEMMA 5.16. *For any* $s \in \widehat{\mathbb{R}}$, *the functions* $\{\phi_n(s;p,q)\}$ *defined in* (5.26) *satisfy*

$$\sum_{n=0}^{\infty}\phi_n(s;p,q) = \sum_{n=0}^{p-1}\phi_n(s;p,q) = 1 \qquad \forall p, q \in \mathbb{Z}^+. \tag{5.33}$$

Proof. Recall that we have proven in (5.27) that $\phi_n(s;p,q) = 0$ for all $n \geq p$. Thus the series (5.33) contains only finite sums. This fact, together with (5.30)–(5.32), implies that

$$\sum_{n=0}^{\infty}\phi_n(s;p,q) = \sum_{n=0}^{p-1}\phi_n(s;p,q) = -\Phi_p(s;p,q) + \Phi_0(s;p,q) = -0 + 1 = 1. \quad \square$$

Combining the results in Lemmas 5.14 and 5.16, we obtain the following crucial identity (5.34). It is worth noting that the discovery of this identity (5.34) is highly nontrivial and becomes the key to proving Proposition 5.11.

LEMMA 5.17. *For any* $s \in \widehat{\mathbb{R}}$, *we have*

$$\sum_{i=1}^{\infty} \nu_{i,p}^{(s)} \nu_{i,q+1}^{(s)} + \sum_{i=1}^{\infty} \nu_{i,p+1}^{(s)} \nu_{i,q}^{(s)} = 2\theta_p^{(s)} \theta_q^{(s)} \qquad \forall p, q \in \mathbb{Z}^+,\ p \equiv q+1 \pmod 2. \tag{5.34}$$

Note the series in (5.34) *is actually* finite *sums, since* $\nu_{i,j}^{(s)} = 0$ *when* $i > j$ *by definition* (5.15).

Proof. Observing that p and q are symmetric in (5.34) and $p \equiv q+1 \pmod 2$, we assume, without loss of generality, that p is odd and q is even (otherwise, we can simply exchange p and q), and denote

$$p = 2\widehat{p} - 1, \quad q = 2\widehat{q} \qquad \text{with } \ \widehat{p}, \widehat{q} \in \mathbb{Z}^+.$$

According to definition (5.15), $\nu_{i,p}^{(s)} = 0$ if i is even, and $\nu_{i,q}^{(s)} = 0$ if i is odd. Thus

$$\sum_{i=1}^{\infty} \nu_{i,p}^{(s)} \nu_{i,q+1}^{(s)} + \sum_{i=1}^{\infty} \nu_{i,p+1}^{(s)} \nu_{i,q}^{(s)} = \sum_{i=1}^{\infty} \nu_{2i-1,2\widehat{p}-1}^{(s)} \nu_{2i-1,2\widehat{q}+1}^{(s)} + \sum_{i=1}^{\infty} \nu_{2i,2\widehat{p}}^{(s)} \nu_{2i,2\widehat{q}}^{(s)}. \tag{5.35}$$

It follows from Lemmas 5.14 and 5.16 that

$$\begin{aligned} &\sum_{i=1}^{\infty} \nu_{2i-1,2\widehat{p}-1}^{(s)} \nu_{2i-1,2\widehat{q}+1}^{(s)} + \sum_{i=1}^{\infty} \nu_{2i,2\widehat{p}}^{(s)} \nu_{2i,2\widehat{q}}^{(s)} \\ &\quad \overset{(5.28)}{=} 2\sum_{i=1}^{\infty} \theta_{2\widehat{p}-1}^{(s)} \theta_{2\widehat{q}}^{(s)} \phi_{i-1}(s; \widehat{p}, \widehat{q}) \overset{(5.33)}{=} 2\theta_{2\widehat{p}-1}^{(s)} \theta_{2\widehat{q}}^{(s)} = 2\theta_p^{(s)} \theta_q^{(s)}, \end{aligned}$$

which along with (5.35) yields (5.34). The proof is completed. ∎

5.5. Proof of Proposition 5.11.

Proof. Note that $\gamma_{p,q} = \gamma_{q,p}$, so that p and q are symmetric in (5.17). Without loss of generality, we assume in the following proof that $p \le q$. The proof is divided into three parts.

(i) Prove (5.17) for $p \not\equiv q \pmod 2$. In this case, $(-1)^{p-1} + (-1)^{q-1} = 0$, and thus $\gamma_{p-1,q-1}^{(s)} = 0$. By (5.15), we know for any given $i \in \mathbb{Z}^+$ that either $\nu_{i,p}^{(s)} = 0$ or $\nu_{i,q}^{(s)} = 0$. Therefore, $\sum_{i=1}^{\min\{p,q\}} \nu_{i,p}^{(s)} \nu_{i,q}^{(s)} = 0 = -\gamma_{p-1,q-1}$.

(ii) Prove (5.17) for the special case $q \ge p = 1$ and $p \equiv q \pmod 2$, namely,

$$\sum_{i=1}^{\min\{1,q\}} \nu_{i,1}^{(s)} \nu_{i,q}^{(s)} = -\gamma_{0,q-1}^{(s)} \qquad \forall q \ge 1, \quad p \equiv q \pmod 2, \tag{5.36}$$

where the left-hand side term is $\nu_{1,1}^{(s)} \nu_{1,q}^{(s)}$, and the right-hand side term is $-\gamma_{0,q-1}^{(s)} = 2\theta_0^{(s)} \theta_q^{(s)}$ by (5.13). Using (5.24) and noting q is odd in this case, we have $\nu_{1,1}^{(s)} \nu_{1,q}^{(s)} = 4\theta_1^{(s)} \theta_q^{(s)} = 2\theta_0^{(s)} \theta_q^{(s)}$. Hence (5.36) holds.

(iii) Prove (5.17) for $q \ge p > 1$ and $p \equiv q \pmod 2$. Since $\nu_{i,j} = 0$ when $i > j$, we can rewrite

$$\sum_{i=1}^{\min\{p,q\}} \nu_{i,p}^{(s)} \nu_{i,q}^{(s)} = \sum_{i=1}^{\infty} \nu_{i,p}^{(s)} \nu_{i,q}^{(s)}. \tag{5.37}$$

We first give the following technical splittings (note all the series below are actually finite sums):

$$\begin{aligned}\sum_{i=1}^{\infty}\nu_{i,p}^{(s)}\nu_{i,q}^{(s)} &= \sum_{k=0}^{p-1}(-1)^k\sum_{i=1}^{\infty}\nu_{i,p-k}^{(s)}\nu_{i,q+k}^{(s)} - \sum_{k=1}^{p-1}(-1)^k\sum_{i=1}^{\infty}\nu_{i,p-k}^{(s)}\nu_{i,q+k}^{(s)}\\ &= \sum_{k=1}^{p}(-1)^{k-1}\sum_{i=1}^{\infty}\nu_{i,p-k+1}^{(s)}\nu_{i,q+k-1}^{(s)} + \sum_{k=1}^{p-1}(-1)^{k-1}\sum_{i=1}^{\infty}\nu_{i,p-k}^{(s)}\nu_{i,q+k}^{(s)}\\ &= \sum_{k=1}^{p-1}(-1)^{k-1}\left(\sum_{i=1}^{\infty}\nu_{i,p-k+1}^{(s)}\nu_{i,q+k-1}^{(s)} + \sum_{i=1}^{\infty}\nu_{i,p-k}^{(s)}\nu_{i,q+k}^{(s)}\right) + (-1)^{p-1}\sum_{i=1}^{\infty}\nu_{i,1}^{(s)}\nu_{i,q+p-1}^{(s)}.\end{aligned}$$

Applying Lemma 5.17 with $\tilde{p} = p-k \in \mathbb{Z}^+$, $\tilde{q} = q+k-1 \in \mathbb{Z}^+$, and $\tilde{p} \equiv \tilde{q}+1 \pmod 2$, we get

$$\sum_{i=1}^{\infty}\nu_{i,p-k}^{(s)}\nu_{i,q+k}^{(s)} + \sum_{i=1}^{\infty}\nu_{i,p-k+1}^{(s)}\nu_{i,q+k-1}^{(s)} = 2\theta_{p-k}^{(s)}\theta_{q+k-1}^{(s)}.$$

Therefore,

$$\begin{aligned}\sum_{i=1}^{\infty}\nu_{i,p}^{(s)}\nu_{i,q}^{(s)} &= \sum_{k=1}^{p-1}(-1)^{k-1}\left(2\theta_{p-k}^{(s)}\theta_{q+k-1}^{(s)}\right) + (-1)^{p-1}\sum_{i=1}^{\infty}\nu_{i,1}^{(s)}\nu_{i,q+p-1}^{(s)}\\ &\overset{(5.36)}{=} \sum_{k=1}^{p-1}(-1)^{k-1}\left(2\theta_{p-k}^{(s)}\theta_{q+k-1}^{(s)}\right) + (-1)^{p-1}\left(2\theta_0^{(s)}\theta_{q+p-1}^{(s)}\right)\\ &= 2\sum_{k=1}^{p}(-1)^{k-1}\theta_{p-k}^{(s)}\theta_{q+k-1}^{(s)} = 2\sum_{j=0}^{p-1}(-1)^{p-j-1}\theta_j^{(s)}\theta_{p+q-j-1}^{(s)}\\ &= 2(-1)^{p-1}\sum_{j=0}^{p-1}(-1)^j\theta_j^{(s)}\theta_{p+q-j-1}^{(s)} = -\gamma_{p-1,q-1}^{(s)}.\end{aligned}$$

This together with (5.37) completes the proof of Proposition 5.11. ∎

6. Numerical results. This section gives a few numerical examples to confirm the theoretical results.

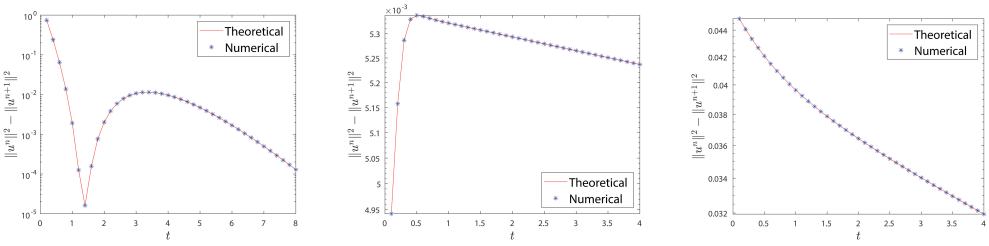
Example 6.1. The first example considers a linear seminegative system from [34]:

$$\frac{\mathrm{d}}{\mathrm{d}t}u = Lu, \quad u = u(t) \in L^2([0,T],\mathbb{R}^3), \quad L = -\begin{pmatrix}1 & 2 & 2\\ 0 & 1 & 2\\ 0 & 0 & 1\end{pmatrix}.$$

The (s,s) diagonal Padé approximations with $s=3$ and $s=4$ are used to solve this system with an arbitrarily chosen initial condition $u(0) = (0.9134, 0.2785, 0.5469)^\top$ up to $t=8$. In order to verify the convergence, we run the simulations with different time stepsizes $\tau \in \{1.6, 0.8, 0.4, 0.2\}$. The l^2-errors in the numerical solutions and the energy dissipation accuracy (see Remark 5.7 for the definition) are listed in Table 1. We observe the convergence rate of $2s$ for the (s,s) diagonal Padé approximation, as expected. We also plot the energy dissipation magnitudes $\|u^n\|^2 - \left\|u^{n+1}\right\|^2$ over time in Figure 1(a). One can observe that $\|u^n\|^2 - \left\|u^{n+1}\right\|^2$ is always positive, which indicates the energy decay property as expected from the unconditionally strong stability in Theorem 5.2. Moreover, the numerical energy dissipation magnitudes agree well with the theoretical ones, which further confirms the correctness of our energy identity (5.6).

TABLE 1
The l^2-errors and energy dissipation accuracy ΔE at $t = 8$, and the corresponding convergence rates for the (s, s) diagonal Padé approximations.

τ	$s = 3$				$s = 4$			
	l^2 error	order	ΔE	order	l^2 error	order	ΔE	order
1.6	3.56e-6	–	1.35e-7	–	2.77e-8	–	1.07e-9	–
0.8	5.25e-8	6.09	1.98e-9	6.09	1.12e-10	7.96	4.34e-12	7.95
0.4	8.07e-10	6.02	3.05e-11	6.02	4.39e-13	7.99	1.71e-14	7.99
0.2	1.26e-11	6.01	4.74e-13	6.01	1.64e-15	8.07	6.36e-17	8.07



(a) Example 6.1. (b) Example 6.2. (c) Example 6.3.

FIG. 1. *Numerical energy dissipation magnitudes and the theoretical ones given by the energy identity* (5.6).

Example 6.2. This example investigates the following seminegative ODE system:

$$\frac{\mathrm{d}}{\mathrm{d}t}u = Lu, \quad u = u(t) \in L^2([0,T], \mathbb{R}^{2N_d}), \quad L = \frac{1}{\Delta x}\begin{pmatrix} \mathbf{L}_1 & \sqrt{3}\mathbf{L}_1 \\ \sqrt{3}(2\mathbf{I}_{N_d} - \mathbf{L}_2) & -3\mathbf{L}_2 \end{pmatrix} \tag{6.1}$$

with

$$\mathbf{L}_1 := \begin{pmatrix} -1 & & & 1 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix}, \qquad \mathbf{L}_2 := \begin{pmatrix} 1 & & & 1 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 1 \end{pmatrix}. \tag{6.2}$$

This system arises from the piecewise linear ($\mathbb{P}^1$-based) discontinuous Galerkin discretization [6] of the linear convection PDE $\psi_t + \psi_x = 0$ in the spatial domain $[0, 1]$ with the uniform mesh of $N_d = 20$ cells (i.e., $\Delta x = 1/N_d = 0.05$) and periodic boundary conditions. The initial solution is taken as $\psi(x, 0) = \sin(2\pi x)$. We solve the semidiscrete ODE system (6.1) in time up to $t = 4$ by using the $(2, 2)$ diagonal Padé approximation. Due to its unconditional strong stability (Theorem 5.2), a large time stepsize $\tau = 0.1$ is used and works robustly. The energy dissipation information shown in Figure 1(b) further validates our theoretical energy laws (5.6) and stability analysis.

Example 6.3. In this example, we study the following seminegative ODE system:

$$\frac{\mathrm{d}}{\mathrm{d}t}u = Lu, \quad u = u(t) \in L^2([0,T], \mathbb{R}^{N_d}), \quad L = \frac{1}{\Delta x^3}\mathbf{L}_1\mathbf{L}_1^\top\mathbf{L}_1^\top \tag{6.3}$$

with the matrix $\mathbf{L}_1$ defined by (6.2). This system comes from the piecewise constant ($\mathbb{P}^0$-based) local discontinuous Galerkin discretization of the dispersion PDE $\psi_t + \psi_{xxx} = 0$ in the spatial domain $[0, 1]$ with the uniform mesh of $N_d = 20$ cells (i.e., $\Delta x = 1/N_d = 0.05$) and periodic boundary conditions. The initial solution is taken as

$\psi(x,0)=\cos(2\pi x)$. We solve the semidiscrete ODE system (6.3) in time up to $t=4$ by using the $(2,2)$ diagonal Padé approximation. The unconditional stability proved in Theorem 5.2 allows us to use a much larger time stepsize $\tau=0.1$, which is not restricted by the normal CFL condition $\Delta t \le C\Delta x^3$ for an explicit time discretization such system (6.3). Figure 1(c) displays the energy dissipation behavior, which is consistent with our theoretical analysis.

7. Conclusions. We have established a systematic theoretical framework to derive the discrete energy laws of general implicit and explicit RK methods for linear seminegative systems. The framework is motivated by a discrete analogue of integration by parts technique and a series expansion of the continuous energy law. The established discrete energy laws show a precise characterization on whether and how the energy dissipates in the RK discretization, thereby giving stability criteria of RK methods. We have also found a unified discrete energy law for all the diagonal Padé approximations, based on analytically constructing the Cholesky type decomposition of a class of symmetric matrices, whose structure is highly complicated. The discovery of the unified energy law and the proof of the decomposition are very nontrivial. For the diagonal Padé approximations, our analyses have bridged the continuous and discrete energy laws, enhancing our understanding of their intrinsic mechanisms. We have provided several specific examples of implicit methods to illustrate the discrete energy laws. A few numerical examples have also been given to confirm the theoretical properties. In this paper, we have developed new analysis techniques, with construction of technical combinatorial identities and the theory of hypergeometric series, which were rarely used in previous RK stability analyses and may motivate future developments in this field. For the future work, we would be interested in extending the proposed framework to multistep methods as a complement of the energy-based analysis of the G-stability [39, Chapter V.6]. The analysis may involve additional difficulty on handling the inner product terms $\langle L^i u^{n+p}, L^j u^{n+q}\rangle$, which depend on both powers of L and solutions at different steps.

Appendix A. Proof of Lemma 2.2.

Proof. For any $v\in V$, we have

$$\begin{aligned}
\sum_{i=0}^{N}\sum_{j=0}^{N}\gamma_{i,j}\tau^{i+j+1}\left[L^i v, L^j v\right] &= -\sum_{i=0}^{N}\sum_{j=0}^{N}\left(\sum_{k=0}^{N}\mu_{k,i}d_k\mu_{k,j}\right)\tau^{i+j+1}\left[L^i v, L^j v\right]\\
&= -\sum_{k=0}^{N} d_k\tau\left(\sum_{i=0}^{N}\sum_{j=0}^{N}\mu_{k,i}\mu_{k,j}\tau^{i+j}\left[L^i v, L^j v\right]\right)\\
&= -\sum_{k=0}^{N} d_k\tau\left[\sum_{i=k}^{N}\tau^i\mu_{k,i}L^i v, \sum_{j=k}^{N}\mu_{k,j}\tau^j L^j v\right]\\
&= -\sum_{k=0}^{N} d_k\tau\left[\!\!\left[\sum_{i=k}^{N}\tau^j\mu_{k,j}L^j v\right]\!\!\right]^2 = -\sum_{k=0}^{N} d_k\tau^{2k+1}\left[\!\!\left[L^k\left(\sum_{j=k}^{N}\mu_{k,j}(\tau L)^{j-k}\right)v\right]\!\!\right]^2. \quad \square
\end{aligned}$$

Appendix B. Lemma B.1 and its proof. Lemma B.1 gives the Cholesky decomposition of a specific matrix $(\widehat{\gamma}_{i,j})$, which is used in the proof of Theorem 2.1 for the energy law of the continuous problem. Note that this special matrix $(\widehat{\gamma}_{i,j})$ is not the matrix $(\gamma_{i,j})$ in (3.10) for the stability analysis of a general RK method.

LEMMA B.1. *Let* $\widehat{\mathbf{\Upsilon}} = (\widehat{\gamma}_{i,j})_{i,j=0}^{N}$ *and* $\widehat{\gamma}_{i,j} = -\frac{1}{i!j!(i+j+1)}$. *Then it holds that*

$$\widehat{\mathbf{\Upsilon}} = -\widehat{\mathbf{U}}^{\top}\widehat{\mathbf{D}}\widehat{\mathbf{U}},$$

where $\widehat{\mathbf{D}} = \mathrm{diag}(\{\widehat{d}_k\}_{k=0}^{N})$ *is a diagonal matrix with* $\widehat{d}_k$ *defined in* (2.3), *and* $\widehat{\mathbf{U}} = (\widehat{\mu}_{k,j})_{k,j=0}^{N}$ *is an upper triangular matrix with* $\widehat{\mu}_{k,j}$ *defined in* (2.3) *for* $j \geq k$ *and* $\widehat{\mu}_{k,j} = 0$ *for* $j < k$.

Proof. Observe that $\widehat{\mathbf{\Upsilon}} = -\mathbf{D}_0\mathbf{H}\mathbf{D}_0$, where $\mathbf{D}_0 = \mathrm{diag}(\{d_i\}_{i=0}^{N})$ with $d_i := 1/i!$, and $\mathbf{H} = (h_{i,j})_{i,j=0}^{N}$ is the Hilbert matrix with $h_{i,j} := 1/(i+j+1)$. The Cholesky decomposition of the Hilbert matrix $\mathbf{H}$ gives $\mathbf{H} = \mathbf{U}_H^{\top}\mathbf{D}_H\mathbf{U}_H$, where the formulae of $\mathbf{U}_H$ and $\mathbf{D}_H$ were given in [17, section 2] and also studied in [15, Lemma 2]. Therefore, we have

$$\widehat{\mathbf{\Upsilon}} = -\mathbf{D}_0\mathbf{U}_H^{\top}\mathbf{D}_H\mathbf{U}_H\mathbf{D}_0 = -\left(\mathbf{D}_0^{-1}\mathbf{U}_H\mathbf{D}_0\right)^{\top}\left(\mathbf{D}_0\mathbf{D}_H\mathbf{D}_0\right)\left(\mathbf{D}_0^{-1}\mathbf{U}_H\mathbf{D}_0\right).$$

Taking $\widehat{\mathbf{U}} = \mathbf{D}_0^{-1}\mathbf{U}_H\mathbf{D}_0$ and $\widehat{\mathbf{D}} = \mathbf{D}_0\mathbf{D}_H\mathbf{D}_0$ with the formulae of $\mathbf{U}_H$ and $\mathbf{D}_H$ from [17, section 2], we obtain (2.3) and complete the proof. ∎

Appendix C. Proof of Theorem 2.1.

Proof. Because $\sum_{i=0}^{\infty}\left\|\frac{1}{i!}(\tau L)^i u(t^n)\right\| \leq \sum_{i=0}^{\infty}\frac{1}{i!}(\tau\|L\|)^i\|u(t^n)\| \leq e^{\tau\|L\|}\|u(t^n)\| \leq e^{T\|L\|}\|u(t^n)\| < \infty$, we known that the series $\sum_{i=0}^{\infty}\frac{1}{i!}(\tau L)^i u(t^n)$ converges. This implies that $v(t^n+\tau) := \sum_{i=0}^{\infty}\frac{1}{i!}(\tau L)^i u(t^n)$ is well-defined. We can verify that $\frac{d}{d\tau}v = Lv$. By the uniqueness of the solution to (1.1), we get $u(t^n+\tau) = v(t^n+\tau) = \sum_{i=0}^{\infty}\frac{1}{i!}(\tau L)^i u(t^n)$. Define $u_N(t^n+\tau) := \sum_{i=0}^{N}\frac{1}{i!}(\tau L)^i u(t^n)$. As $N\to\infty$, we have $\|u_N - u\| \to 0$ and thus $[\![u_N]\!] \to [\![u]\!]$. It then follows from (1.4) that

(C.1)
$$\left\|u(t^{n+1})\right\|^2 - \|u(t^n)\|^2 = -\int_0^{\tau}[\![u(t^n+\widehat{\tau})]\!]^2\,\mathrm{d}\widehat{\tau} = -\int_0^{\tau}\lim_{N\to\infty}[\![u_N(t^n+\widehat{\tau})]\!]^2\,\mathrm{d}\widehat{\tau}.$$

Using the inequality $[\![u]\!]^2 \leq 2\|L\|\|u\|^2$, we deduce that

(C.2)
$$\begin{aligned}[\![u_N(t^n+\widehat{\tau})]\!]^2 &\leq 2\|L\|\|u_N(t^n+\widehat{\tau})\|^2\\ &\leq 2\|L\|\left(\sum_{i=0}^{N}\frac{1}{i!}\widehat{\tau}^i\|L\|^i\right)^2\|u(t^n)\|^2 \leq 2\|L\|e^{2\widehat{\tau}\|L\|}\|u(t^n)\|^2.\end{aligned}$$

Thanks to the *dominated convergence theorem*, the estimate (C.2) along with $\int_0^{\tau}2\|L\|e^{2\widehat{\tau}\|L\|}\|u(t^n)\|^2\,\mathrm{d}\widehat{\tau} = (e^{2\tau\|L\|}-1)\|u(t^n)\|^2 < \infty$ implies $\int_0^{\tau}\lim_{N\to\infty}[\![u_N(t^n+\widehat{\tau})]\!]^2\,\mathrm{d}\widehat{\tau} = \lim_{N\to\infty}\int_0^{\tau}[\![u_N(t^n+\widehat{\tau})]\!]^2\,\mathrm{d}\widehat{\tau}$. Combining it with (C.1) gives

(C.3)
$$\left\|u(t^{n+1})\right\|^2 - \|u(t^n)\|^2 = -\lim_{N\to\infty}\int_0^{\tau}[\![u_N(t^n+\widehat{\tau})]\!]^2\,\mathrm{d}\widehat{\tau}.$$

On the other hand, we can reformulate the integration (C.3) as follows:

$$\begin{aligned}\int_0^{\tau}[\![u_N(t^n+\widehat{\tau})]\!]^2\,d\widehat{\tau} &= \int_0^{\tau}\left[\sum_{i=0}^{N}\frac{1}{i!}(\widehat{\tau}L)^i u(t^n), \sum_{j=0}^{N}\frac{1}{j!}(\widehat{\tau}L)^j u(t^n)\right]d\widehat{\tau}\\ &= \sum_{i,j=0}^{N}\left(\int_0^{\tau}\frac{\widehat{\tau}^{i+j}}{i!j!}\mathrm{d}\widehat{\tau}\right)\left[L^i u(t^n), L^j u(t^n)\right]\\ &= \sum_{i,j=0}^{N}\frac{\widehat{\tau}^{i+j+1}}{i!j!(i+j+1)}\left[L^i u(t^n), L^j u(t^n)\right] = \sum_{k=0}^{N}\widehat{d}_k\tau^{2k+1}\left[\!\left[L^k\widehat{u}_N^{(k)}\right]\!\right]^2,\end{aligned} \tag{C.4}$$

where the last equality follows from Lemmas 2.2 and B.1, $\widehat{u}_N^{(k)} := \sum_{j=k}^{N} \widehat{\mu}_{k,j}(\tau L)^{j-k} u(t^n)$, and $\widehat{d}_k$ and $\widehat{\mu}_{k,j}$ are defined in (2.3). Hence, by combining (C.4) with (C.3), we obtain

$$\begin{aligned} \left\|u(t^{n+1})\right\|^2 - \|u(t^n)\|^2 &= -\lim_{N\to\infty} \sum_{k=0}^{N} \widehat{d}_k \tau^{2k+1} \left[\!\left[L^k \widehat{u}_N^{(k)} \right]\!\right]^2 \\ &= -\lim_{N\to\infty} \sum_{k=0}^{\infty} \widehat{d}_k \tau^{2k+1} \left[\!\left[L^k \widehat{u}_N^{(k)} \right]\!\right]^2 1_{\{0\le k\le N\}}, \end{aligned} \tag{C.5}$$

where $1_{\{\cdot\}}$ is the indicator function. Note that

$$\widehat{d}_k \tau^{2k+1} \left[\!\left[L^k \widehat{u}_N^{(k)} \right]\!\right]^2 1_{\{0\le k\le N\}} \le 2\tau\|L\|\|u(t^n)\|^2 \left(\sum_{j=k}^{N} \sqrt{\widehat{d}_k}\widehat{\mu}_{k,j}(\tau\|L\|)^j \right)^2 =: \mathcal{B}_k. \tag{C.6}$$

The upper bound $\mathcal{B}_k$ satisfies

$$\sum_{k=0}^{\infty} \mathcal{B}_k \le \|u(t^n)\|^2 \left(e^{2\tau\|L\|} - 1 \right) < \infty, \tag{C.7}$$

because for any integer $M \ge N$,

$$\begin{aligned} \sum_{k=0}^{M} \mathcal{B}_k &\le 2\tau\|L\|\|u(t^n)\|^2 \sum_{k=0}^{M} \left(\sum_{j=k}^{M} \sqrt{\widehat{d}_k}\widehat{\mu}_{k,j}(\tau\|L\|)^j \right)^2 \\ &= 2\tau\|L\|\|u(t^n)\|^2 \sum_{i=0}^{M}\sum_{j=0}^{M} \frac{(\tau\|L\|)^{i+j}}{i!j!(i+j+1)} = 2\|u(t^n)\|^2 \int_0^{\tau\|L\|} \left(\sum_{i=0}^{M} \frac{x^i}{i!} \right)^2 \mathrm{d}x \\ &\le 2\|u(t^n)\|^2 \int_0^{\tau\|L\|} e^{2x} \mathrm{d}x = \|u(t^n)\|^2 \left(e^{2\tau\|L\|} - 1 \right), \end{aligned}$$

where we have used Lemma B.1 in the first equality. Due to (C.6) and (C.7), we can again invoke the dominated convergence theorem to exchange the limit and the infinite summation in (C.5) to obtain

$$\left\|u(t^{n+1})\right\|^2 - \|u(t^n)\|^2 = -\sum_{k=0}^{\infty} \lim_{N\to\infty} \widehat{d}_k \tau^{2k+1} \left[\!\left[L^k \widehat{u}_N^{(k)} \right]\!\right]^2 1_{\{0\le k\le N\}} = -\sum_{k=0}^{\infty} \widehat{d}_k \tau^{2k+1} \left[\!\left[L^k \widehat{u}^{(k)} \right]\!\right]^2,$$

which completes the proof. ∎

Appendix D. Proof of Lemma 5.4.

Proof. When $i > j$ or $i = j$, the identity is obviously true. In the following, we only focus on the case of $i < j$. Define

$$\begin{aligned} a_\ell &:= \binom{s-\ell}{j-\ell}^{-1} \binom{2s-\ell}{j-i-\ell} \binom{i+j+1}{\ell} (-1)^\ell \\ &= \frac{(j-\ell)!(s-j)!(2s-\ell)!(i+j+1)!}{(s-\ell)!(j-i-\ell)!(2s-j+i)!\ell!(i+j+1-\ell)!} (-1)^\ell. \end{aligned}$$

Then we have

$$a_0 = \frac{j!(s-j)!(2s)!}{s!(j-i)!(2s-j+i)!}, \qquad \frac{a_\ell}{a_0} = \prod_{k=0}^{\ell-1} \frac{a_{k+1}}{a_k} = \prod_{k=0}^{\ell-1} \left(\frac{(k-j+i)(k-s)(k-i-j-1)}{(k-2s)(k-j)} \frac{1}{k+1} \right).$$

Using the rising factorial notation (5.19), one can reformulate the sum in Lemma 5.4 as

(D.1)
$$\sum_{\ell=0}^{j-i} a_\ell = a_0 \sum_{\ell=0}^{j-i} \frac{(i-j)_\ell (-s)_\ell (-i-j-1)_\ell}{(-2s)_\ell (-j)_\ell} \frac{1}{\ell!} = a_0 \sum_{\ell=0}^{\infty} \frac{(i-j)_\ell (-s)_\ell (-i-j-1)_\ell}{(-2s)_\ell (-j)_\ell} \frac{1}{\ell!},$$

where we have used the fact $(i-j)_\ell = 0$ for $\ell > j-i$ and $\ell \in \mathbb{N}$. By using the notation ${}_3F_2(\,)$ from the theory of generalized hypergeometric functions [41], the above series can also be represented as

$$\sum_{\ell=0}^{\infty} \frac{(i-j)_\ell (-s)_\ell (-i-j-1)_\ell}{(-2s)_\ell (-j)_\ell} \frac{1}{\ell!} = {}_3F_2\left(\begin{matrix} i-j, & -s, & -i-j-1 \\ & -2s, & -j \end{matrix} \right)$$
$$= {}_3F_2\left(\begin{matrix} -n, & c, & 2c+2d+n-1 \\ & 2c, & c+d \end{matrix} \right)$$

with $n := j-i \in \mathbb{N}$, $c := -s$, and $d := s-j$. We use Watson's formula [41] for such hypergeometric series:

(D.2)
$${}_3F_2\left(\begin{matrix} -n, & c, & 2c+2d+n-1 \\ & 2c, & c+d \end{matrix} \right) = \begin{cases} \dfrac{n!\Gamma(c+\frac{1}{2}n)\Gamma(d+\frac{1}{2}n)\Gamma(2c)\Gamma(c+d)}{(\frac{1}{2}n)!\Gamma(c+d+\frac{1}{2}n)\Gamma(2c+n)\Gamma(c)\Gamma(d)} & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd.} \end{cases}$$

If $n = j-i$ is even, define $m := \frac{j-i}{2} = \frac{1}{2}n \in \mathbb{N}$. Note that the singularity in (D.2) is removable, because

(D.3)
$$\frac{\Gamma(x+m)}{\Gamma(x)} = \frac{(x+m-1)\Gamma(x+m-1)}{\Gamma(x)} = \cdots = \prod_{\ell=0}^{m-1} (x+\ell) \neq 0, \qquad x = c, d, c+d,$$

(D.4)
$$\frac{\Gamma(2c)}{\Gamma(2c+n)} = \frac{\Gamma(2c)}{(2c+n-1)\Gamma(2c+n-1)} = \cdots = \prod_{\ell=0}^{n-1} \frac{1}{2c+\ell} = \frac{(2s-j+i)!}{(2s)!} > 0,$$

where the formula $\Gamma(x+1) = x\Gamma(x)$ has been used repeatedly. It follows from (D.3) that

(D.5)
$$\frac{\Gamma(c+\frac{1}{2}n)}{\Gamma(c)} = \prod_{\ell=0}^{m-1} (-s+\ell) = (-1)^m \frac{s!}{(s-m)!} = (-1)^m \frac{s!}{(s-\frac{j-i}{2})!},$$

(D.6)
$$\frac{\Gamma(d+\frac{1}{2}n)}{\Gamma(d)} = \prod_{\ell=0}^{m-1} (s-j+\ell) = (s-j)\frac{(s-j+m-1)!}{(s-j)!} = (s-j)\frac{(s-1-\frac{i+j}{2})!}{(s-j)!},$$

(D.7)
$$\frac{\Gamma(c+d)}{\Gamma(c+d+\frac{1}{2}n)} = (-1)^m \prod_{\ell=0}^{m-1} (j-\ell)^{-1} = (-1)^m \frac{(j-m)!}{j!} = (-1)^m \frac{(\frac{i+j}{2})!}{j!}.$$

Substituting (D.4)–(D.7) into (D.2) and combining (D.1) with (D.2), we obtain for $i \equiv j \pmod 2$ that

$$\begin{aligned}\sum_{\ell=0}^{j-i} a_\ell &= a_0 \frac{n!}{(\frac{1}{2}n)!}\frac{s!}{(s-\frac{j-i}{2})!}(s-j)\frac{(s-1-\frac{i+j}{2})!}{(s-j)!}\frac{(2s-j+i)!}{(2s)!}\frac{(\frac{i+j}{2})!}{j!}\\ &= \frac{(s-1-\frac{i+j}{2})!(\frac{i+j}{2})!}{(s-\frac{j-i}{2})!(\frac{j-i}{2})!}(s-j),\end{aligned}$$

which completes the proof. ∎

Appendix E. Proof of Lemma 5.12.

Proof. By the definition of the Pochhammer symbol, one can deduce that

$$\begin{aligned}(x+n)! &= x!(x+1)(x+2)\cdots(x+n) = x!(x+1)_n,\\ (x)_n &= \left(\prod_{0\le 2i\le n-1}(x+2i)\right)\cdot\left(\prod_{0\le 2i+1\le n-1}(x+2i+1)\right)\\ &= x(x+2)\cdots\left(x+2\left\lceil\frac{n}{2}\right\rceil-2\right)\cdot(x+1)(x+3)\cdots\left(x+2\left\lfloor\frac{n}{2}\right\rfloor-1\right)\\ &= 2^{\lceil\frac{n}{2}\rceil}\left(\frac{x}{2}\right)_{\lceil\frac{n}{2}\rceil}\cdot 2^{\lfloor\frac{n}{2}\rfloor}\left(\frac{x+1}{2}\right)_{\lfloor\frac{n}{2}\rfloor} = 2^n\left(\frac{x}{2}\right)_{\lceil\frac{n}{2}\rceil}\left(\frac{x+1}{2}\right)_{\lfloor\frac{n}{2}\rfloor},\\ \frac{(x+i)!}{(x-j)!} &= (x-j+1)(x-j+2)\cdots(x-1)x\cdot(x+1)\cdots(x+i)\\ &= (-1)^j(-x)_j(x+1)_i.\end{aligned}$$

∎

Appendix F. Proof of Lemma 5.13.

Proof. If $i > j$, then by definition (5.15) we know that $\nu^{(s)}_{2i,2j} = \nu^{(s)}_{2i-1,2j-1} = 0$. On the other hand, when $i > j$, we have $(-j)_i = 0$ and $(1-j)_{i-1} = 0$, which imply the right-hand sides of (5.23) and (5.24) are both zero. Hence the identities (5.23) and (5.24) are true for $i > j$. In the following, we focus on the nontrivial case that $i \le j$.

Proof of (5.23) *for* $i \le j$. We observe that

$$\nu^{(s)}_{2i,2j} = \frac{s!}{(2s)!}\frac{2\sqrt{4i-1}}{(2i+2j)!}\frac{(2s+2i-2j)!(s-i-j)!(i+j)!}{(s-2j)!(s-j+i)!(j-i)!} = \frac{s!}{(2s)!}\frac{2\sqrt{4i-1}}{(s-2j)!}\Pi_1\Pi_2$$

with

$$\begin{aligned}\Pi_1 &:= \frac{(s-i-j)!}{(s-j+i)!}(2s-2j+2i)! \overset{(5.20)}{=} \frac{(s-j-i)!}{(s-j+i)!}(2s-2j)!(2s-2j+1)_{2i}\\ &\overset{(5.21)}{=} \frac{(s-j-i)!}{(s-j+i)!}(2s-2j)!2^{2i}\left(s-j+\frac{1}{2}\right)_i(s-j+1)_i\\ &\overset{(5.22)}{=} \frac{1}{(-1)^i(j-s)_i(s-j+1)_i}(2s-2j)!2^{2i}\left(s-j+\frac{1}{2}\right)_i(s-j+1)_i\\ &= \frac{(2s-2j)!2^{2i}\left(s-j+\frac{1}{2}\right)_i}{(j-s)_i(-1)^i},\end{aligned}$$

and

$$\Pi_2 := \frac{(j+i)!}{(j-i)!}\frac{1}{(2i+2j)!} \overset{(5.22)}{=} \frac{(-1)^i(-j)_i(j+1)_i}{(2i+2j)!} \overset{(5.20)}{=} \frac{(-1)^i(-j)_i(j+1)_i}{(2j)!(2j+1)_{2i}}$$
$$\overset{(5.21)}{=} \frac{(-1)^i(-j)_i(j+1)_i}{(2j)!2^{2i}\left(j+\frac{1}{2}\right)_i(j+1)_i} = \frac{(-1)^i(-j)_i}{2^{2i}\left(j+\frac{1}{2}\right)_i}\frac{1}{(2j)!}.$$

It follows that

$$\Pi_1\Pi_2 = \frac{(2s-2j)!}{(2j)!}\frac{\left(s-j+\frac{1}{2}\right)_i(-j)_i}{(j-s)_i\left(j+\frac{1}{2}\right)_i}.$$

Therefore, we obtain

$$\nu^{(s)}_{2i,2j} = 2\sqrt{4i-1}\frac{s!}{(2s)!}\frac{(2s-2j)!}{(2j)!(s-2j)!}\frac{\left(s-j+\frac{1}{2}\right)_i(-j)_i}{(j-s)_i\left(j+\frac{1}{2}\right)_i}$$
$$= 2\sqrt{4i-1}\theta^{(s)}_{2j}\frac{\left(s-j+\frac{1}{2}\right)_i(-j)_i}{(j-s)_i\left(j+\frac{1}{2}\right)_i},$$

which yields (5.23).

Proof of (5.24) *for* $i \le j$. We observe that

$$\nu^{(s)}_{2i-1,2j-1} = \frac{s!}{(2s)!}\frac{2\sqrt{4i-3}}{(2i+2j-2)!}\frac{(2s-2j+2i)!}{(s-j+i)!}\frac{(s-i-j+1)!}{(s-2j+1)!}\frac{(j+i-1)!}{(j-i)!}$$
$$= \frac{s!}{(2s)!}\frac{2\sqrt{4i-3}}{(s-2j+1)!}\Pi_3\Pi_4$$

with

$$\Pi_3 := \frac{(s-i-j+1)!}{(s-j+i)!}(2s-2j+2i)! \overset{(5.20)}{=} \frac{(s-i-j+1)!}{(s-j+i)!}(2s-2j+1)!(2s-2j+2)_{2i-1}$$
$$\overset{(5.21)}{=} \frac{(s-i-j+1)!}{(s-j+i)!}(2s-2j+1)!2^{2i-1}(s-j+1)_i\left(s-j+\frac{3}{2}\right)_{i-1}$$
$$\overset{(5.22)}{=} \frac{(2s-2j+1)!}{(-1)^{i-1}(j-s)_{i-1}(s-j+1)_i}2^{2i-1}(s-j+1)_i\left(s-j+\frac{3}{2}\right)_{i-1}$$
$$= \frac{(2s-2j+1)!2^{2i-1}\left(s+\frac{3}{2}-j\right)_{i-1}}{(-1)^{i-1}(j-s)_{i-1}},$$

and

$$\Pi_4 := \frac{(j+i-1)!}{(j-i)!}\frac{1}{(2j+2i-2)!} \overset{(5.22)}{=} \frac{(-1)^i(-j)_i(j+1)_{i-1}}{(2j+2i-2)!} = \frac{(-1)^{i-1}(1-j)_{i-1}(j)_i}{(2j+2i-2)!}$$
$$\overset{(5.20)}{=} \frac{(-1)^{i-1}(1-j)_{i-1}(j)_i}{(2j-1)!(2j)_{2i-1}} \overset{(5.21)}{=} \frac{(-1)^{i-1}(1-j)_{i-1}(j)_i}{(2j-1)!2^{2i-1}(j)_i\left(j+\frac{1}{2}\right)_{i-1}}$$
$$= \frac{(-1)^{i-1}(1-j)_{i-1}}{2^{2i-1}\left(j+\frac{1}{2}\right)_{i-1}}\frac{1}{(2j-1)!}.$$

It follows that

$$\Pi_3\Pi_4 = \frac{(2s-2j+1)!}{(2j-1)!}\frac{\left(s+\frac{3}{2}-j\right)_{i-1}}{(j-s)_{i-1}}\frac{(1-j)_{i-1}}{\left(j+\frac{1}{2}\right)_{i-1}}.$$

Therefore, we complete the proof by noting

$$\begin{aligned}\nu^{(s)}_{2i-1,2j-1} &= 2\sqrt{4i-3}\frac{s!}{(2s)!}\frac{(2s-2j+1)!}{(2j-1)!(s-2j+1)!}\frac{\left(s+\frac{3}{2}-j\right)_{i-1}(1-j)_{i-1}}{(j-s)_{i-1}\left(j+\frac{1}{2}\right)_{i-1}}\\ &= 2\sqrt{4i-3}\theta^{(s)}_{2j-1}\frac{\left(s+\frac{3}{2}-j\right)_{i-1}(1-j)_{i-1}}{(j-s)_{i-1}\left(j+\frac{1}{2}\right)_{i-1}}.\end{aligned}$$

∎

REFERENCES

[1] G. Birkhoff and R. S. Varga, *Discretization errors for well-set Cauchy problems*, J. Math. Phys., 44 (1965).

[2] C. Bresten, S. Gottlieb, Z. Grant, D. Higgs, D. Ketcheson, and A. Németh, *Explicit strong stability preserving multistep Runge–Kutta methods*, Math. Comp., 86 (2017), pp. 747–769.

[3] E. Burman, A. Ern, and M. A. Fernández, *Explicit Runge–Kutta schemes and finite elements with symmetric stabilization for first-order linear PDE systems*, SIAM J. Numer. Anal., 48 (2010), pp. 2019–2042.

[4] K. Burrage and J. C. Butcher, *Stability criteria for implicit Runge–Kutta methods*, SIAM J. Numer. Anal., 16 (1979), pp. 46–57.

[5] J. C. Butcher, *Numerical Methods for Ordinary Differential Equations*, John Wiley & Sons, New York, 2016.

[6] B. Cockburn and C.-W. Shu, *The local discontinuous Galerkin method for time-dependent convection-diffusion systems*, SIAM J. Numer. Anal., 35 (1998), pp. 2440–2463.

[7] G. G. Dahlquist, *A special stability problem for linear multistep methods*, BIT, 3 (1963), pp. 27–43.

[8] E. Deriaz, *Stability conditions for the numerical solution of convection-dominated problems with skew-symmetric discretizations*, SIAM J. Numer. Anal., 50 (2012), pp. 1058–1085.

[9] D. Drake, J. Gopalakrishnan, J. Schöberl, and C. Wintersteiger, *Convergence analysis of some tent-based schemes for linear hyperbolic systems*, Math. Comp., 91 (2022), pp. 699–733.

[10] B. L. Ehle, *A-stable methods and Padé approximations to the exponential*, SIAM J. Math. Anal., 4 (1973), pp. 671–680.

[11] B. L. Ehle and Z. Picel, *Two-parameter, arbitrary order, exponential approximations for stiff equations*, Math. Comp., 29 (1975), pp. 501–511.

[12] J. Gopalakrishnan and Z. Sun, *Stability of Structure-Aware Taylor Methods for Tents*, preprint, arXiv:2203.05176, 2022.

[13] S. Gottlieb, C.-W. Shu, and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Rev., 43 (2001), pp. 89–112.

[14] E. Hairer, G. Bader, and C. Lubich, *On the stability of semi-implicit methods for ordinary differential equations*, BIT, 22 (1982), pp. 211–232.

[15] E. Hairer and G. Wanner, *Algebraically stable and implementable Runge–Kutta methods of high order*, SIAM J. Numer. Anal., 18 (1981), pp. 1098–1108.

[16] I. Higueras, *Monotonicity for Runge–Kutta methods: Inner product norms*, J. Sci. Comput., 24 (2005), pp. 97–117.

[17] S. Hitotumatu, *Cholesky decomposition of the Hilbert matrix*, Jpn. J. Appl. Math., 5 (1988), pp. 135–144.

[18] A. Iserles, *A First Course in the Numerical Analysis of Differential Equations*, Cambridge University Press, Cambridge, 2009.

[19] L. Isherwood, Z. J. Grant, and S. Gottlieb, *Strong stability preserving integrating factor Runge–Kutta methods*, SIAM J. Numer. Anal., 56 (2018), pp. 3276–3307.

[20] D. I. Ketcheson, *Relaxation Runge–Kutta methods: Conservation and stability for inner-product norms*, SIAM J. Numer. Anal., 57 (2019), pp. 2850–2870.

[21] D. I. Ketcheson, S. Gottlieb, and C. B. Macdonald, *Strong stability preserving two-step Runge–Kutta methods*, SIAM J. Numer. Anal., 49 (2011), pp. 2618–2639.

[22] J. Kraaijevanger and M. Spijker, *Algebraic stability and error propagation in Runge–Kutta methods*, Appl. Numer. Math., 5 (1989), pp. 71–87.

[23] J. F. B. M. Kraaijevanger, *Contractivity of Runge–Kutta methods*, BIT, Numer. Math., 31 (1991), pp. 482–528.

[24] D. Levy and E. Tadmor, *From semidiscrete to fully discrete: Stability of Runge–Kutta schemes by the energy method*, SIAM Rev., 40 (1998), pp. 40–73.
[25] J. V. Neumann, *Eine spektraltheorie für allgemeine operatoren eines unitären raumes*, Math. Nachr., 4 (1950), pp. 258–281.
[26] P. Öffner, J. Glaubitz, and H. Ranocha, *Analysis of artificial dissipation of explicit and implicit time-integration methods*, Int. J. Numer. Anal. Model., 17 (2020).
[27] M. Qin and M. Zhang, *Symplectic Runge–Kutta algorithms for Hamiltonian systems*, J. Comput. Math., 10 (1992), pp. 205–215.
[28] H. Ranocha, *On strong stability of explicit Runge–Kutta methods for nonlinear semibounded operators*, IMA J. Numer. Anal., 41 (2021), pp. 654–682.
[29] H. Ranocha and D. I. Ketcheson, *Energy stability of explicit Runge–Kutta methods for nonautonomous or nonlinear problems*, SIAM J. Numer. Anal., 58 (2020), pp. 3382–3405.
[30] H. Ranocha and D. I. Ketcheson, *Relaxation Runge–Kutta methods for Hamiltonian problems*, J. Sci. Comput., 84 (2020), pp. 1–27.
[31] H. Ranocha and P. Öffner, *L_2 stability of explicit Runge–Kutta schemes*, J. Sci. Comput., (2018), pp. 1–17.
[32] H. Ranocha, M. Sayyari, L. Dalcin, M. Parsani, and D. I. Ketcheson, *Relaxation Runge–Kutta methods: Fully discrete explicit entropy-stable schemes for the compressible Euler and Navier–Stokes equations*, SIAM J. Sci. Comput., 42 (2020), pp. A612–A638.
[33] M. Spijker, *Contractivity in the numerical solution of initial value problems*, Numer. Math., 42 (1983), pp. 271–290.
[34] Z. Sun and C.-W. Shu, *Stability of the fourth order Runge–Kutta method for time-dependent partial differential equations*, Ann. Math. Sci. Appl., 2 (2017), pp. 255–284.
[35] Z. Sun and C.-W. Shu, *Strong stability of explicit Runge–Kutta time discretizations*, SIAM J. Numer. Anal., 57 (2019), pp. 1158–1182.
[36] Z. Sun and C.-W. Shu, *Enforcing strong stability of explicit Runge–Kutta methods with superviscosity*, Commun. Appl. Math. Comput., 3 (2021), pp. 671–700.
[37] E. Tadmor, *From semidiscrete to fully discrete: Stability of Runge–Kutta schemes by the energy method*. II, in Collected Lectures on the Preservation of Stability Under Discretization, Proc. Appl. Math. 109, Lecture Notes from Colorado State University Conference, Fort Collins, CO, 2001 D. Estep and S. Tavener, eds., SIAM, Philadelphia, 2002, pp. 25–49.
[38] H. Wang, C.-W. Shu, and Q. Zhang, *Stability and error estimates of local discontinuous Galerkin methods with implicit-explicit time-marching for advection-diffusion problems*, SIAM J. Numer. Anal., 53 (2015), pp. 206–227.
[39] G. Wanner and E. Hairer, *Solving Ordinary Differential Equations* II*: Stiff and Difterential-Algebraic Problems*, 1st ed., Springer, Berlin, 1991.
[40] G. Wanner, E. Hairer, and S. P. Nørsett, *Order stars and stability theorems*, BIT, 18 (1978), pp. 475–489.
[41] G. N. Watson, *A note on generalized hypergeometric series*, Proc. Lond. Math. Soc., 23 (1925), pp. xiii–xv.
[42] Y. Xu, X. Meng, C.-W. Shu, and Q. Zhang, *Superconvergence analysis of the Runge–Kutta discontinuous Galerkin methods for a linear hyperbolic equation*, J. Sci. Comput., 84 (2020), pp. 1–40.
[43] Y. Xu, Q. Zhang, C.-W. Shu, and H. Wang, *The L^2-norm stability analysis of Runge–Kutta discontinuous Galerkin methods for linear hyperbolic equations*, SIAM J. Numer. Anal., 57 (2019), pp. 1574–1601.
[44] Q. Zhang and C.-W. Shu, *Stability analysis and a priori error estimates of the third order explicit Runge–Kutta discontinuous Galerkin method for scalar conservation laws*, SIAM J. Numer. Anal., 48 (2010), pp. 1038–1063.